HUMBOLDT-UNIVERSITÄT ZU BERLIN
INSTITUT FÜR BIBLIOTHEKS- UND INFORMATIONSWISSENSCHAFT

BERLINER HANDREICHUNGEN
ZUR BIBLIOTHEKS- UND
INFORMATIONSWISSENSCHAFT

HEFT 278

GESCHÄFTSMODELLE FÜR DIE DIGITALE LANGZEITARCHIVIERUNG

DAS BEISPIEL FORSCHUNGSDATEN

VON
DANIEL BEUCKE

Geschäftsmodelle für die digitale Langzeitarchivierung

das Beispiel Forschungsdaten

von
Daniel Beucke

Berliner Handreichungen zur
Bibliotheks- und Informationswissenschaft

Begründet von Peter Zahn
Herausgegeben von
Konrad Umlauf
Humboldt-Universität zu Berlin

Heft 278

Beucke, Daniel

Geschäftsmodelle für die digitale Langzeitarchivierung – das Beispiel Forschungsdaten / von Daniel Beucke. - Berlin : Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, 2010. – 65 S.: graph. Darst. - (Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft ; 278)

ISSN 14 38-76 62

Abstract:

Forschungsdaten liegen zunehmend in digitaler Form vor bzw. werden ausschließlich digital produziert. Die Produktion ist zum einen sehr aufwendig und Kostenintensiv, zum anderen können sie Daten nicht ein zweites Mal erhoben werden. Daraus ergeben sich für Gedächtnisorganisationen wie Bibliotheken, Archive und Museen sowie auch für die Industrie bezogen auf die Möglichkeit der dauerhaften Verfügbarkeit dieser Daten neue Herausforderungen: Sie müssen Strategien und Geschäftsmodelle entwickeln, um den mit dieser neuen Situation verbundenen Anforderungen gerecht zu werden.
In der vorliegenden Arbeit wird untersucht, wie ein Geschäftsmodell für die Langzeitarchivierung (LZA) von digitalen Forschungsdaten beschaffen sein sollte und welche Komponenten für eine erfolgreiche Umsetzung der LZA notwendig sind.
Abschließend werden konkreten Empfehlungen für Geschäftsmodelle für die digitale Langzeitarchivierung von Forschungsdaten gegeben.

Diese Veröffentlichung geht zurück auf eine Master-Arbeit (unter dem damaligen Namen des Verfassers Daniel Metje) im postgradualen Fernstudiengang Master of Arts (Library and Information Science) an der Humboldt-Universität zu Berlin.

Online-Version: http://edoc.hu-berlin.de/series/berliner-handreichungen/2010-278

Inhalt

Abbildungsverzeichnis

Tabellenverzeichnis

1 Einleitung

Daten liegen zunehmend in digitaler Form vor bzw. werden nur noch digital produziert. Dies hat zur Folge, dass Gedächtnisorganisationen wie Bibliotheken, Archive und Museen sowie auch die Industrie bezogen auf die Möglichkeit der dauerhaften Verfügbarkeit dieser Daten vor neuen Herausforderungen stehen: Sie müssen Strategien und Geschäftsmodelle entwickeln, um den mit dieser neuen Situation verbundenen Anforderungen gerecht zu werden.

Der Begriff Langzeitarchivierung wird in der vorliegenden Arbeit im Sinne des Open Archival Information Systems (OAIS) Reference Model (ISO 14721:2003) (CCSDS Secretariat 2002) für eine gesicherte, langfristige Archivierung und Bereitstellung interpretierbarer und zitierfähiger Daten verwendet: „The act of maintaining information, in a correct and Independently Understandable form, over the Long Term.“ (CCSDS Secretariat 2002, S. 1-6) Dabei versteht man unter ‚Long Term‘ in diesem Kontext "A period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community, on the information being held in a repository. This period extends into the indefinite future." (CCSDS Secretariat 2002, S. 1-6) Die im Rahmen der Langzeitarchivierung entwickelten Strategien müssen somit auf unbestimmte Zeit jede technische und kulturelle Veränderung berücksichtigen, denen digitale Objekte unterworfen sein können.

Digitale Langzeitarchivierung ist dabei mehr als nur die Aufbewahrung des digitalen Objektes und seines Inhalts: Immer muss darauf geachtet werden, dass auch die relevanten Eigenschaften (Significant Properties) erhalten werden. Selbstverständlich stellt auch die Vertrauenswürdigkeit und Beständigkeit des digitalen Archivs einen bedeutenden Aspekt von Langzeitarchivierung dar.

Digitale Objekte lassen sich nach Thibodeau (Thibodeau 2002, S. 6 ff.) in drei verschiedene Ebenen einteilen. Es wird unterschieden zwischen dem physischen Objekt, dem logischen Objekt und schließlich dem konzeptuellen Objekt. Als physisches Objekt werden die binären Zeichen auf einem Datenträger verstanden (Bits und Bytes). Die Erhaltung dieses Datenflusses ist Grundvoraussetzung für spätere Erhaltungs-Strategien. Als logisches Objekt wird die Folge und der Zusammenhang der einzelnen Bits gesehen, welche aus dem digitalen Objekt ein Datei-Format erkennen lassen. Das konzeptuelle Objekt beinhaltet die gesamte Funktionalität, die dem Benutzer durch die Verwendung der passenden Hard- und Software zur Verfügung steht. Langzeitarchivierung schließt alle drei Ebenen ein.

Das Problem der Langzeitarchivierung[1] wird schnell als ein nur Technisches gesehen, das mit der richtigen Hard- und Software gelöst werden kann. Langzeitarchivierung geht jedoch weit über die Bitstream-Preservation hinaus, die lediglich den Ursprung – die Bits und Bytes – des digitalen Objektes sichert.

Zurzeit existieren mehrere Forschungsprojekte auf dem Gebiet der Langzeitarchivierung. Auf nationaler Ebene beschäftigt sich beispielsweise nestor – das Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen in Deutschland[2] mit Fragestellungen der Langzeitarchivierung auf verschiedenen Gebieten. Auf europäischer Ebene vermittelt das Projekt DigitalPreservationEurope (DPE)[3] das Thema Langzeitarchivierung in den europäischen Mitgliedstaaten. Anfang des Jahres richtete DPE einen eigenen Channel[4] auf der Internetplattform YouTube ein, um mit einem Animationsfilm[5] über Probleme der Langzeitarchivierung von digitalen Daten zu informieren.

Das digitale Zeitalter hat auch für die Forschung eine Reihe von Änderungen mit sich gebracht. So liegen heutzutage viele Forschungsdaten[6] ausschließlich in digitaler Form vor. Dies lässt sich wie folgt begründen: „Research findings in digital form can be easily moved around, duplicated, handed to others, worked on with new tools, merged with other data, divided up in new ways, stored in vast volumes and manipulated by supercomputers if their nature so demands.“ (RIN 2008, S. 7)

Zu der Problematik der Langzeitarchivierung von Forschungsdaten liegen bereits erste Studien und Ergebnisse aus Forschungsprojekten[7] vor. Der Nutzen der Langzeitarchivierung von Forschungsdaten wird in der Studie „Keeping research data save“ sehr treffend auf den Punkt gebracht: „The benefits [...] in preserving research data extend from the fact that the discoveries of the future rely on the work of the past.“ (Beagrie 2008, S. 16) Weitere Vorteile, die Daten in einem Archiv zur Verfügung zu stellen, sind:

- Sicherung von Investitionen in die Forschung

---

[1] In der vorliegenden Arbeit werden im Rahmen der Langzeitarchivierung verwendete Fachbegriffe als vorausgesetzt angesehen.

[2] http://www.langzeitarchivierung.de [20.05.2009]. Alle weiteren Links sind ebenfalls an diesem Tag auf Erreichbarkeit hin überprüft wurden.

[3] http://www.digitalpreservationeurope.eu/

[4] http://www.youtube.com/user/wepreserve

[5] http://www.youtube.com/watch?v=pbBa6Oam7-w&feature=channel_page

[6] Mit dem Thema Forschungsdaten setzt sich das zweite Kapitel näher auseinander.

[7] Vgl. u.a. Allianz-Initiative Digitale Information 2008, Beagrie/Chruszcz/Lavoie 2008, Lyon 2007, Klump 2008b, RIN 2008.

- Sicherung der Daten für zukünftige Fragestellungen bzw. Überprüfung von Forschungsergebnissen
- Vermarktung von Forschung
- Unterstützung von Forschungs- und Lernprozessen

Im Rahmen einer Umfrage an zwei Wissenschaftsstandorten (Universität Göttingen und ETH Zürich) wurde der Umgang mit anfallenden Forschungsdaten erhoben und ausgewertet (vgl. Kapitel 2.1). Die Ergebnisse dieser Umfrage dienen in der vorliegenden Arbeit als Grundlage und verdeutlichen die Probleme und Aktivitäten im Bereich der Langzeitarchivierung an zwei verschiedenen Wissenschaftsstandorten. Der Standort Göttingen wird in der vorliegenden Arbeit gelegentlich als Praxisbeispiel herangezogen, um komplexe Darstellungen zu konkretisieren. Verschiedene Göttinger Einrichtungen beschäftigen sich bereits mit dem Thema der Langzeitarchivierung im Allgemeinen sowie im Besonderen mit der Langzeitarchivierung von Forschungsdaten. Hierzu zählt auch das unten beschriebene Projekt KoLaWiss. Zudem ist die Universität Göttingen Projektpartner im europäischen Projekt PARSE.Insight - INSIGHT into issues of Permanent Access to the Records of Science in Europe[8], dessen Ziel die Ausarbeitung einer Roadmap und von Empfehlungen zur Unterstützung der e-Infrastruktur zur Langzeitarchivierung von Forschungsdaten ist.

Geschäftsmodelle[9] sind für alle an der Langzeitarchivierung von Forschungsdaten Beteiligten von entscheidender Bedeutung, damit eine funktionierende Infrastruktur am Standort finanziert und etabliert werden kann. In der vorliegenden Arbeit werden die Schwierigkeiten bei der Entwicklung solcher Geschäftsmodelle aufgezeigt.

Im Jahr 2007 hat die Deutsche Forschungsgemeinschaft (DFG) diese Dringlichkeit erkannt und eine Aktionslinie zum Thema „Entwicklung von Organisations- und Geschäftsmodellen“ initiiert (DFG 2007). Die Ausschreibung von Forschungsprojekten bezog sich auf elektronische Dokumente in Form von, E-Journals, Webseiten, PDF-Dokumenten usw. der Sondersammelgebietsbibliotheken (SSG-Bibliotheken), Digitalisate aus DFG-geförderten Digitalisierungsprojekten und Forschungsdaten aus DFG-geförderten Forschungsvorhaben. In den beiden Letztgenannten wurde in München und Göttingen Förderung bewilligt. Hauptaufgabe dieser Projekte war die Entwicklung von Organisations- und Geschäftsmodellen. Zudem galt es, Fragestellungen zu den Themen Technik, Recht und Kosten zu beantworten. Das Münchner Pro-

[8] http://www.parse-insight.eu/

[9] Eine genaue Definition zu Geschäftsmodellen für die Langzeitarchivierung wird in dem Kapitel 3 der Arbeit gegeben.

jekt[10] beschäftigte sich mit Geschäftsmodellen für Digitalisate aus DFG-geförderten Digitalisierungsprojekten; das Göttinger Projekt KoLaWiss[11] widmete sich inhaltlich der Materialart Forschungsdaten. Ein Arbeitspaket „Entwicklung von Organisations- und Geschäftsmodellen“ war in beiden Projekten vorhanden. In den Projekten wurden zwei unterschiedliche Herangehensweisen bei der Bearbeitung der Fragestellung verfolgt: So wurde im Rahmen des KoLaWiss-Projekts ein Bottom-Up Ansatz gewählt und zunächst die anfallenden Aufgaben, die beteiligten Partner und deren Rolle in der Langzeitarchivierung von Forschungsdaten am Beispiel des Göttinger Standortes betrachtet, um auf dieser Grundlage ein generisches Modell zu entwickeln. Im Münchner Projekt wurde hingegen eine Top-Down Strategie angewendet, bei der sich der Problematik zunächst auf einer abstrakten Ebene genähert wurde, indem unterschiedliche Modelle und Prozesse beispielsweise bezogen auf Nutzer und Erträge betrachtet und definiert wurden. An Hand von kopal und Portico werden Überlegungen zu Geschäftsmodellen für eine Langzeitarchivierung vorgestellt und deren Vor- und Nachteile untersucht. Im Mittelpunkt steht dabei die Überprüfung, inwieweit diese Modelle sich für eine Langzeitarchivierung von Forschungsdaten eignen.

Zielsetzung der Arbeit und methodisches Vorgehen

Die Archivierung von und der Zugriff auf Forschungsdaten ist also für die Wissenschaft von grundlegender Wichtigkeit. Gleichzeitig liegen immer mehr Forschungsdaten nur noch in digitaler Form vor.

In der vorliegenden Arbeit wird untersucht, wie ein Geschäftsmodell für die Langzeitarchivierung von Forschungsdaten beschaffen sein sollte und welche Komponenten für eine erfolgreiche Umsetzung notwendig sind. Angrenzende Themen, wie z.B. die technische Umsetzung oder rechtliche Fragen der Langzeitarchivierung von Forschungsdaten werden in der Arbeit nicht behandelt.

Zunächst werden die Besonderheiten von Forschungsdaten und ihrer Langzeitarchivierung betrachtet. Dazu werden Ergebnisse einer Umfrage an den Wissenschaftsstandorten Göttingen und Zürich zur Langzeitarchivierung hinzugezogen, da diese einen detaillierten Überblick über die Probleme geben, die bei der Langzeitarchivierung entstehen können. Darstellt werden zudem die Ergebnisse des Projektes KoLaWiss, die in Göttingen auf zwei Workshops präsentiert wurden. Abschließend wird im zweiten Kapitel ein Überblick über bereits vorhandene Empfehlungen und Policies zur Langzeit-

---

[10] Projektpartner sind die Bayrische Staatsbibliothek und die Universität der Bundeswehr München.

[11] Projektpartner sind die Gesellschaft für wissenschaftliche Datenverarbeitung mbH, die Staats- und Universitätsbibliothek Göttingen, der Geschäftsbereich Informationstechnologie und die Medizinische Informatik der Universitätsmedizin Göttingen.

archivierung von Forschungsdaten gegeben, die Auswirkungen auf die Umsetzung von Strategien und Organisationsmodellen im Bereich der Langzeitarchivierung von Forschungsdaten haben werden.

Anschließend werden die an der Langzeitarchivierung beteiligten Akteure am Beispiel des Wissenschaftsstandortes Göttingen vorgestellt und verdeutlicht, welche Rollen an einem Wissenschaftsstandort von den dort befindlichen Institutionen und Einrichtungen übernommen werden müssen. Unter Berücksichtigung der verschiedenen Komponenten, wie z.B. Kostenberechnung, Service-Level-Agreements und vertragliche Regelungen sowie Finanzierung, werden Grundvoraussetzungen für die Entwicklung eines Geschäftsmodells in Kapitel drei aufgezeigt. Bereits vorhandene Geschäftsmodelle werden exemplarisch vorgestellt.

Im letzten Kapitel erfolgt eine Zusammenstellung der in der Arbeit erörterten Aspekte bezogen auf die Entwicklung von Geschäftsmodellen in der Langzeitarchivierung von Forschungsdaten. Offene Fragen werden diskutiert. Die Arbeit schließt mit konkreten Empfehlungen für Geschäftsmodelle für die digitale Langzeitarchivierung von Forschungsdaten.

2 Besonderheiten bei der Langzeitarchivierung von Forschungsdaten

Der Begriff Forschungsdaten wird im Folgenden synonym verwendet mit Rohdaten oder Primärdaten. In der deutschsprachigen Literatur werden alle diese Begriffe benutzt, was zu Verwirrungen führen kann (vgl. dazu Klump 2008b, S. 15-54). In der englischsprachigen Literatur werden die Begriffe Research Data oder Scientific Data benutzt.

Forschungsdaten in digitaler Form spielen für die Wissenschaft eine immer wichtigere Rolle. Sie entstehen bei Experimenten und bei Beobachtungen spezifischer Ereignisse und umfassen Modelle und Simulationen (vgl. RIN 2008, S. 7). Die Daten sind ein Produkt sich ständig ändernder Fragestellungen und können unterschiedliche Formen der Transformation im Forschungsprozess durchleben. Die ursprünglichen Rohdaten werden normalerweise nicht veröffentlicht. Bei einer Publikation solcher Daten handelt es sich größtenteils um weiterverarbeitete oder transformierte Daten (vgl. RIN 2008, S. 7f).

Die Langzeitarchivierung von Forschungsdaten wird durch verschiedene Faktoren beeinträchtigt. Aufgrund der Vielzahl der wissenschaftlichen Fachdisziplinen liegen die Forschungsdaten in sehr heterogenen Dateiformaten vor. Diese Tatsache erschwert die Langzeitarchivierung dieser Daten erheblich. Dies liegt unter anderem auch daran, dass in den Disziplinen kaum Richtlinien (Policies) für die Erstellung und für die zu verwendenden Dateiformate von Forschungsdaten vorliegen. Auch die nicht vorhandenen oder nur unvollständigen Metadaten erschweren die Langzeitarchivierung der Daten. „Diese Faktoren stellen für die digitale Langzeitarchivierung von wissenschaftlichen Primärdaten eine größere Herausforderung dar, als die Datenmenge, auch wenn diese in einzelnen Fällen sehr groß sein kann.“ (Klump 2008b, S. 15-54) Besonders große Datenmengen fallen beispielsweise in der Hochenergiephysik und der Astronomie an.

Oft liegen zu den digitalen Objekten nur unvollständige, falsche oder gar keine Metadaten vor. Der Grund für die schlechte Qualität der vorhandenen Metadaten und die nicht vorhandene Format-Standardisierung liegt zumeist darin begründet, dass kaum ausreichende zeitliche oder finanziellen Ressourcen vorhanden sind und dass die Produzenten der Daten – die Forscher – es nicht als ihre Aufgabe erachten, die weitere Nutzung und die Archivierung der Daten zu unterstützen und zu verlangen (vgl. Klump et al. 2006, S. 80). Das liegt unter anderem auch daran, dass die „Währung“ der Wissenschaft die zitierbare Veröffentlichung darstellt. Erst durch eine Publikation erhalten die Forscher ihr Renommée, da dadurch die Arbeit des Wissenschaftlers rezipiert werden kann. In den seltensten Fällen besteht für die Wissenschaftler die Notwendigkeit, ihre Forschungsdaten zu publizieren. Der Großteil der Wissenschaftler sieht in der Veröffentlichung der Daten deshalb keinen An-

reiz und nimmt diesen Mehraufwand nur selten in Kauf, da dadurch keine Reputation erfolgt (vgl. Klump 2008b, 15-54). Lediglich die Hoffnung auf Zusammenarbeit und die Unterstützung der Gutachter sind Motive, die Daten aufzubereiten und zu veröffentlichen (vgl. RIN, S 8).

Erste Ansätze, auch die eigentlichen Forschungsdaten zu veröffentlichen, werden von europäischen Forschungsbibliotheken und Technischen Informationszentren[12] vorgestellt, die sich mit dem Ziel zusammengeschlossen haben, den Zugang zu Forschungsdaten im Internet zu verbessern (vgl. TIB 2009). Dafür sollen die Forschungsdaten mit Persistent Identifier versehen werden, so dass sie als eindeutige, unabhängige und zitierbare wissenschaftliche Objekte behandelt werden können. Die Technische Informationsbibliothek in Hannover (TIB) vergibt dazu den Forschungsdaten Digital Object Identifier (DOI), um eine dauerhafte (persistente) Adressierung der digitalen Objekte zu gewährleisten. Seit 2005 hat die TIB mit diesem Verfahren ca. 600.000 Forschungsdatensätze in ihren Katalog aufnehmen können, die im Internet auffindbar und damit für jeden frei nutzbar sind (vgl. TIB 2009).

Der offene Zugang zu Forschungsdaten (auch Open Data genannt) wird mittlerweile durch verschiedene Projekte und Institutionen unterstützt und gefördert. Im September 2008 fand zu diesem Thema ein Offener Helmholtz Workshop „Umgang mit Forschungsdaten“ in Bremen statt, bei dem renommierte Wissenschaftler über den freien Zugang zu Forschungsdaten referierten und im Anschluss diskutierten. Das DFG-Projekt „Publication and Citation of Scientific Primary Data“ (STD-DOI)[13] beschäftigt sich mit der Zitierbarkeit von wissenschaftlichen Forschungsdaten, damit die Forschungsdaten im Internet auffindbar und nutzbar sind. Das durch die Helmholtz-Gesellschaft unterstützte Projekt „Publishing Network for Geoscientific & Environmental Data (PANGAEA)“[14] ist eine Plattform, in der wissenschaftliche Forschungsdaten aus dem Bereich der Geowissenschaft zu finden und frei verfügbar sind. Dies sind nur einige Beispiele, wie Forschungsdaten im wissenschaftlichen Kontext eine immer wichtigere Bedeutung erhalten.

Um die Langzeitarchivierung von Forschungsdaten zu erleichtern bzw. überhaupt zu ermöglichen, muss ein Paradigmenwechsel in der Wissenschaftskultur stattfinden. Den Wissenschaftlern und damit den Produzenten der Forschungsdaten muss verdeutlicht werden, warum das Überführen der Daten in ein Langzeitarchiv notwendig ist.

---

[12] Die Technische Informationsbibliothek Hannover (TIB), die British Library, die Bibliothek der ETH Zürich, das französische Institute for Scientific and Technical Information (INIST), das Technical Information Center aus Dänemark und die TU Delft Bibliothek aus den Niederlanden.

[13] http://www.std-doi.de/ [Stand: 18.05.09].

[14] http://www.pangaea.de/

Erste Ansätze der zuvor genannten Projekte zur Zitierfähigkeit und der damit verbundenen Auffindbarkeit von Forschungsdaten helfen auf diesem Weg der Veränderungen, da die Forscher ein anderes Verhältnis zu ihren Forschungsdaten einnehmen und von vornherein anders mit ihnen arbeiten. Beispielsweise durch die Anreicherung der Daten mit Metadaten erleichtern diese die im Anschluss stattfindende Langzeitarchivierung der Forschungsdaten.

2.1 Umfrage an der Universität Göttingen und der ETH Zürich – Vergleich zweier wissenschaftlicher Standorte

Im Jahr 2007 fand in Kooperation zwischen der SUB Göttingen und der ETH Zürich – mit Unterstützung durch nestor[15] – eine Online-Umfrage statt, die an beiden Wissenschaftsstandorten durchgeführt wurde.

Ziel der an den zwei unterschiedlichen Standorten durchgeführten Umfrage war es, neben der Evaluierung der produzierten und genutzten Daten auch einen möglichen Unterschied im Umgang mit digitalen Daten zwischen einer Volluniversität und einer technisch-naturwissenschaftlichen Universität zu untersuchen (vgl. Metje/Vlaeminck 2008, S. 109). Als weitere Ziele galt es zu klären, inwieweit ein genereller Bedarf an der Langzeitarchivierung von Forschungsdaten besteht und potenzielle Ansprechpartner für weitere Gespräche an den jeweiligen Standorten zu identifizieren. Zudem sollte mit der Umfrage der Bewusstseinstand der Wissenschaftler in Bezug zur Langzeitarchivierung der digitalen Daten ermittelt werden (vgl. Boutsiouci/Metje 2008).

Die Georg-August-Universität Göttingen (Uni Göttingen) gilt mit ihren 13 Fakultäten – einschließlich der Medizin – als eine Volluniversität. Darüber hinaus sind am Wissenschaftsstandort Göttingen weitere Wissenschaftsinstitute, wie die Gesellschaft für wissenschaftliche Datenverarbeitung mbH, das Deutsche Primaten Zentrum und mehrere Max-Planck-Institute angesiedelt.

Die Eidgenössische Technische Hochschule Zürich (ETH Zürich) ist eine technisch-naturwissenschaftliche Universität und hat mit ca. 15.000 Studenten etwas mehr als die Hälfte an eingeschriebenen Studenten wie die Göttinger Universität.

Die beiden beteiligten Institutionen entwickelten einen Online-Survey, der als Evaluierungsinstrument für die in Wissenschaft und Forschung anfallenden Daten genutzt wurde. Ausgangspunkt waren die digitalen Daten aus Forschung (digitale Messdaten und elektronische Auswertungen), Lehre (eLear-

---

[15] nestor – das Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen in Deutschland [http://www.langzeitarchivierung.de]

ning), Bibliothek (elektronische Publikationen) und Verwaltung (digitale Dokumente) (vgl. Boutsiouci/Metje 2008).

Die Ergebnisse der Befragungen ergaben an beiden Standorten, dass ein großes Interesse an der Langzeitarchivierung von digitalen Daten besteht (80%). Jedoch gaben nur 12% der Befragten an, dass erste Planungen im Bezug auf die Archivierung der Daten erfolgt sind. Diese Zahlen sind an beiden Standorten identisch.

Abbildung 1: Besteht in Ihrem Bereich Interesse an der LZA von digitalen Daten? Gibt es bereits Planungen zur LZA?

Die Ergebnisse zeigen eine große Diskrepanz zwischen dem Interesse und den konkreten Planungen an den Standorten. Es wird eine zukünftige Aufgabe sein, dieses Interesse in Strategien zur Langzeitarchivierung umzusetzen und mit den betreffenden Mitarbeitern weiterführende Gespräche zu führen.

Bei der Frage nach Vereinbarungen oder gesetzlichen Vorschriften zur Langzeitarchivierung von Daten gaben an beiden Standorten fast 80% der Befragten an, dass keine solchen Vereinbarungen vorhanden sind. Diejenigen, die das Vorhandensein von gesetzlichen Vorschriften bejahten, verwiesen zumeist auf die von der DFG geforderte gute wissenschaftliche Praxis sowie interne Regelungen der jeweiligen Hochschule.

Abbildung 2: Gibt es in Ihrem Bereich Vereinbarungen oder gesetzliche Vorschriften, welche die Langzeitarchivierung Ihrer Daten regeln?

Diese Werte verdeutlichen, dass an den Wissenschaftsstandorten noch weiterer Informationsbedarf bei den Wissenschaftlern notwendig ist, über die vorliegenden Vorschriften und Richtlinien (vgl. Kapitel 2.3) aufgeklärt zu werden.

Die für die Langzeitarchivierung von digitalen Objekten relevante Eigenschaft der Dateiformate wurde im Rahmen der Umfrage ebenfalls evaluiert. Nicht jedes Dateiformat ist für eine spätere Überführung in ein digitales Langzeitarchiv geeignet. So sind proprietäre Formate oder auch Formate, die nicht ausreichend gut dokumentiert sind, nicht für die Archivierung brauchbar, da im Prozess der Langzeitarchivierung kaum eine Migration oder Emulation auf Grund der fehlenden Informationen möglich ist. Die Antworten bei der Befragung ergaben, dass die Mehrzahl der Daten in den Formaten PDF, JPEG, TIFF und denen der Microsoft-Office-Familie vorhanden sind. 77% der Göttinger und 51% der Züricher Befragten gaben an, dass die genutzten Formate als „allgemein anerkannt" und „gut dokumentiert" charakterisiert werden können.[16]

[16] Es muss jedoch vermutet werden, dass viele Wissenschaftler nicht in ausreichendem Maße darüber informiert sind, was als „allgemein anerkannt" angesehen wird, und was mit „guter Dokumentation" der Formate gemeint ist. Vgl. auch Klump, 2008a, S.16.

Abbildung 3: Liegen die dauerhaft zu archivierenden Daten in einem Standardformat vor, das in ihrem Fach oder allgemein anerkannt und gut dokumentiert ist?

Für eine Langzeitarchivierung von Forschungsdaten ist es notwendig, dass die Daten mit Metadaten angereichert sind. Das sind zum einen die technischen und die für die Langzeitarchivierung relevanten Metadaten, zum anderen die deskriptiven Metadaten, die das Objekt inhaltlich beschreiben. Besonders letztere sind im Nachhinein nur sehr kosten- und zeitintensiv zu ermitteln (vgl. Kapitel 3.2.1.). Daher ist es von besonderer Wichtigkeit, dass diese Metadaten bei dem Entstehen des digitalen Objektes bzw. kurz danach durch den Produzenten hinzugefügt werden. Leider sehen sich zurzeit die Produzenten noch nicht in der Rolle, diese Aufgabe zu übernehmen (vgl. Severins/Hilf 2006, S. 25). Bei der Umfrage an den Standorten Göttingen und Zürich ergaben die Fragen nach der Verwendung von technischen bzw. inhaltlichen Metadaten bei den beiden Standorten Ergebnisse mit leichten Abweichungen in den Antworten. Bezogen auf die Frage nach der Verwendung von technischen Metadaten bei den digitalen Objekten gaben die Befragten des Göttinger Standortes an, dass mehr als drei Viertel der Daten keine technischen Metadaten beinhalten, bei der ETH Zürich waren hiervon nur rund die Hälfte der Daten betroffen. Auf die Frage, ob die betreffenden Daten aussagekräftige administrative und deskriptive Beschreibungen enthalten, die es erlauben, diese zu einem späteren Zeitpunkt sachlich korrekt zu interpretieren, antworteten an beiden Standorten etwa die Hälfte der Befragten, dass diese Daten vorhanden seien.

Abbildung 4: Enthalten die betreffenden digitalen Daten technische bzw. administrative und deskriptive Beschreibungen?

Zum Ende der Befragung wurde evaluiert, durch welche personellen Maßnahmen die Verfügbarkeit der Daten zurzeit sichergestellt wird. An beiden Standorten zeigte sich, dass die Mitarbeiter selbst für eine Verfügbarkeit der Daten verantwortlich sind (ETH 60%, Göttingen 68%). Meist werden beim Ausscheiden der Personen die Daten an den Nachfolger übergeben. Verantwortliche Personen sind nicht klar definiert (ETH 26%, Göttingen 22%). Klare Richtlinien liegen in beiden Fälle kaum vor (ETH 20%, Göttingen 16%).

Abbildung 5: Durch welche Maßnahmen stellen Sie bzw. Ihre Mitarbeiter die Verfügbarkeit Ihrer Daten heute sicher? (Mehrfachnennung möglich)

Bei beiden Standorten kann zusammenfassend festgestellt werden, dass ein großes Interesse an der Langzeitarchivierung von den umfangreich vorliegenden Forschungsdaten vorhanden ist. Jedoch besteht an beiden Standorten großer Handlungsbedarf bei der Entwicklung und Umsetzung von Lösungen, damit eine Langzeitarchivierung für Forschungsdaten erfolgreich sein kann. Große Abweichungen auf Grund der Art der Universitätsstandorte konnten nicht festgestellt werden.

2.2 Ergebnisse aus den KoLaWiss-Workshops

Im Rahmen des DFG-geförderten Forschungsprojekt "Kooperative Langzeitarchivierung für Wissenschaftsstandorte" (KoLaWiss) wurden in Göttingen zwei Workshops zum Thema Langzeitarchivierung von Forschungsdaten organisiert. Der erste Workshop[17], der am 9. Dezember 2008 statt fand, behandelte allgemein das Thema Langzeitarchivierung von Forschungsdaten. So ging es um die bisherigen Projektergebnisse, die einem ausgewählten Personenkreis zur Diskussion gestellt wurden. Die Vorträge beschäftigten sich mit den Themen Datenkategorisierung von Forschungsdaten, Technik, Recht und Kosten. Der letzte Vortrag handelte von Organisations- und Geschäftsmodellen für die Langzeitarchivierung von Forschungsdaten und beschäftigte sich im Speziellen mit den anfallenden Aufgaben bei der Langzeitarchivierung von Forschungsdaten, den beteiligten Institutionen und deren Rollen bei der Organisation (Metje 2008).

Neben dem Vortrag zum Thema Organisations- und Geschäftsmodelle für die Langzeitarchivierung von Forschungsdaten ergab die anschließende Diskussion relevante Ergebnisse in Bezug auf Organisations- und Geschäftsmodelle, die im Folgenden vorgestellt werden:

- Die Akteure werden sich mit der Zeit ändern, und neue Akteure werden hinzu kommen, wie beispielsweise der Datenkurator oder der Preservation Officer, der als zentraler Ansprechpartner am Wissenschaftsstandort agieren könnte.
- Aufgaben werden ggf. an externe Dienstleister vergeben, oder es werden Kooperationen mit der Industrie stärker in den Vordergrund treten. Dabei soll berücksichtigt werden, dass die bereits vorhandenen und strategischen Kompetenzen am Standort verbleiben und weiter ausgebaut werden.
- Öffentliche Einrichtungen sind prädestiniert für den Betrieb von Infrastruktur für die Langzeitarchivierung. Sie genießen großes Vertrauen innerhalb der beteiligten Institutionen und haben eine hohe Bestandssicherheit.

[17] http://kolawiss.uni-goettingen.de/?q=de/node/8

Der zweite KoLaWiss Workshop fand im Januar 2009 statt und trug den Titel „Wie kann die Langzeitarchivierung von Forschungsdaten finanziert werden?“[18]. Bei dieser Veranstaltung standen mögliche Fördermodelle für die Langzeitarchivierung von Forschungsdaten im Vordergrund sowie die Frage, wie eine mögliche Organisationsform der Langzeitarchivierung aussehen könnte. Die Teilnehmer des Workshops kamen aus verschiedenen Einrichtungen und Institutionen. Neben wissenschaftlichen Mitarbeitern aus verschiedenen Projekten waren Mitarbeiter aus Rechenzentren und aus der Wissenschaftsverwaltung, als auch Mitarbeiter von Förderorganisationen und Mitarbeiter verschiedenen Konzerne anwesend.

In einem ersten Block stellten die KoLaWiss-Projektmitarbeiter die erarbeiteten Forschungsergebnisse vor, die sich mit Förderempfehlungen zu den einzelnen Arbeitspaketen beschäftigten. In einem zweiten Block präsentierten verwandte Projekte in diesem Zusammenhang deren Herangehensweise an das Thema Langzeitarchivierung von Forschungsdaten. Abschließend wurde bei einer Podiumsdiskussion[19] über die Möglichkeiten der Finanzierung und Organisation der digitalen Langzeitarchivierung von Forschungsdaten diskutiert.

Die relevanten Ergebnisse, die im Rahmen des zweiten KoLaWiss Workshops zusätzlich zu den Vorträgen in Bezug auf Organisation und Geschäftsmodelle für die Langzeitarchivierung von Forschungsdaten gesammelt werden konnten, werden im Folgenden dargestellt:

- Die diskutierten Geschäftsmodelle sollen individuell für jeden Standort bearbeitet werden. Anhand von Pilotprojekten, die sich mit dem Thema der Geschäftsmodelle beschäftigen, sollen generische Modelle entwickelt werden, die im Folgenden auf andere Standorte angewendet werden können.
- Auf Grund von gesetzlichen Vorgaben für den öffentlichen Dienst ist das vorgestellte Finanzierungsmodelle „Kapitaldeckungsverfahren“[20] für Betriebe des öffentlichen Dienstes nicht ohne Weiteres umsetzbar, da öffentliche Einrichtungen kein Geld auf dem Kapitalmarkt anlegen dürfen.
- Die Universitäten sollen zur Finanzierung der Langzeitarchivierung von Forschungsdaten beitragen. Diese Aufgabe übernehmen sie bereits jetzt, indem sie beispielsweise wissenschaftliche Bibliotheken

[18] http://kolawiss.uni-goettingen.de/?q=de/node/9

[19] Sabine Rey (Moderation; Projektleitung KoLaWiss; Medizinische Informatik der Universitätsmedizin Göttingen), Dr. Jens Klump (GFZ Potsdam), Dr. Hans Pfeiffenberger (AWI), Dr. Heike Neuroth (SUB/MPDL), Prof. Dr. Ulrich Sax (Geschäftsbereich Informationstechnologie der Universitätsmedizin Göttingen) und Dr. Stefan Winkler-Nees (DFG).

[20] Siehe zu den Finanzierungsmodellen Kapitel 3.2.3.

finanzieren, um die wissenschaftliche Literatur- und Informationsversorgung zu gewährleisten.

- Der Ingest der digitalen Daten könnte durch eine ausgelagerte Finanzierung ermöglicht werden, da dieser Schritt ein sehr kostenintensiver im Workflow der Langzeitarchivierung ist[21]. In dem Beispiel eines Workshopsteilnehmers wird die Finanzierung dieses Schrittes an den Datenproduzenten abgegeben, der auch für die Erstellung der deskriptiven Metadaten zuständig ist – ebenfalls ein kostenintensiver Punkt.
- Die Förderorganisationen sollen bei der Projektförderung eine „Datenpauschale" mit in ihr Programm aufnehmen. Analog zu der Publikationspauschale der DFG würden dem Projekt finanzielle Mittel für die Langzeitarchivierung der wissenschaftlichen Forschungsdaten zur Verfügung stehen. Diese Mittel dürfen nach der Bewilligung nicht mehr umgewidmet werden und sollen ausschließlich für die Langzeitarchivierung der anfallenden Daten verwendet werden. Sind weitere Mittel für die Archivierung der Daten notwendig, weil im Verlauf der Forschungen weitere Daten anfallen, die dauerhaft zur Verfügung stehen müssen, dann sollen finanzielle Mittel auch im Nachhinein noch angefordert werden können.
- Ein zentraler Ansprechpartner am Wissenschaftsstandort (Preservation Officer), der für alle Fragen rund um das Thema der digitalen Langzeitarchivierung von Forschungsdaten zuständig ist, wurde als sinnvoll erachtet. Diese Person dient als Ansprechpartner für alle Beteiligten an der Langzeitarchivierung von Forschungsdaten und vermittelt zwischen den Dienstleistungspartnern. Der Preservation Officer kann z.B. bei der Entwicklung und Umsetzung von LZA-Policies am Standort mitarbeiten oder den Fachwissenschaftler bei der Erstellung der Metadaten für die digitalen Objekte unterstützend tätig sein.
- Das System der Fachreferenten an wissenschaftlichen Bibliotheken kann zudem dafür genutzt werden, den einzelnen Wissenschaftler (Produzent der Daten) in fachspezifischen Fragen, die zur Vorbereitung auf die Abgabe der Daten ins Langzeitarchiv dienen, zu unterstützen. Die Bibliotheken haben auf diesem Bereich eine Kompetenz und könnten somit ihr Angebot weiter ausbauen. Ein Vorteil an diesem Modell wäre, dass keine neue Struktur aufgebaut werden müsste, da auf ein funktionierendes, gut organisiertes System zurückgegriffen werden kann.

[21] Siehe dazu im Speziellen Kapitel 3.3.1 Kostenmodelle.

- In der Diskussion wurde zudem darauf hingewiesen, dass die Langzeitarchivierung von Forschungsdaten an öffentliche Institutionen angebunden werden soll, da es sich dabei um vertrauenswürdige Organisationen handelt, die auch über einen gewissen Zeitraum bestehen bleiben.
- Es wurde darüber nachgedacht, die bereits vorhandenen Erfahrungen in Kompetenzzentren zu bündeln und somit sog. „LZA-Knoten" zu schaffen, die die Langzeitarchivierung von Forschungsdaten übernehmen und diese auch als externe Dienstleistung an kleinere Wissenschaftsstandorte und Institutionen anbieten.

2.3 Richtlinien zur Langzeitarchivierung von Forschungsdaten

„Es gilt (...) eine nachhaltige integrierte digitale Forschungsumgebung zu schaffen, in der jeder Forschende von überall in Deutschland auf (...) die relevanten Forschungsprimärdaten zugreifen kann." (Allianz 2008, S. 1)

Mehrere Organisationen beschäftigen sich schon seit längerer Zeit mit Lösungen, um den Zugang und die dauerhafte Erhaltung von digitalen Forschungsdaten zu ermöglichen. Allein in Deutschland liegen die Investitionen für die Gewinnung von digitalen Forschungsdaten in der Größenordnung von mehreren Milliarden Euro pro Jahr (vgl. Allianz-Initiative Digitale Information 2008, S. 5). Dies hat unter anderem dazu geführt, dass sich seit einigen Jahren mehrere Organisationen mit dem Thema des offen Zugangs und der langfristigen Nutzbarkeit von Daten aus öffentlich geförderter Forschung beschäftigen. Daraufhin entstanden in den letzten Jahren verschiedene Policies und Empfehlungen, die sich mit diesem Thema auseinandersetzen und dadurch Veränderungen in der Wissenschaftskultur erwirken wollen.

Vor mehr als zehn Jahren – im Jahr 1998 – gab die Deutsche Forschungsgemeinschaft (DFG) die „Empfehlungen für eine gute wissenschaftliche Praxis" bekannt (DFG 1998). In diesen Empfehlungen fordert die DFG, dass die Institutionen die Forschungsdaten, die als Grundlage einer Publikation dienten, für mindestens zehn Jahre auf sicheren Datenträgern zu speichern haben (vgl. DFG 1998, Empfehlung 7, S. 12). Diese Empfehlung der DFG sagt jedoch nichts über den Zugang zu den Forschungsdaten und über deren Nachnutzbarkeit aus, als dass sie vielmehr der Nachvollziehbarkeit der wissenschaftlichen Arbeit der Wissenschaftler dient.[22] Sie scheint jedoch wenig Anwendung in der Praxis bekommen zu haben, da diese Empfehlung bereits seit über zehn Jahren in Anwendung ist und mit Bedauern festgestellt werden muss, dass die DFG ihre Empfehlungen weder nachprüft noch Sanktionen für das Nicht-Einhalten auferlegt (vgl. Klump 2008b, S. 15-55).

---

[22] Dies ist auch darin begründet, dass es in dieser Zeit eine Reihe von Betrugsfällen gab.

Eine weitere Empfehlung wurde nach einer dreitägigen Konferenz über den offenen Zugang zu wissenschaftlichen Informationen im Oktober 2003 veröffentlicht. Die Ergebnisse dieses Treffen gingen in die „Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen“ ein (Berliner Erklärung 2003). Die Konferenz wurde von der Max-Planck-Gesellschaft initiiert. Die Erklärung unterzeichneten große deutsche Wissenschaftsorganisationen sowie führende internationale Forschungs- und Kultureinrichtungen. Die „Berliner Erklärung“ ist ein Meilenstein in der Open-Access Bewegung und nimmt noch heute eine bedeutende Rolle ein. Die Unterzeichner der Erklärung weisen auf die zunehmende Bedeutung des Internets für die Wissenschaft und das wissenschaftliche Publizieren hin. Das Ziel ist ein offener Zugang zu wissenschaftlichen Daten, zu denen nach der Erklärung auch Forschungsdaten und Metadaten gehören (vgl. Berliner Erklärung 2003). Mit dieser Erklärung legen die Unterzeichner sich auf einen freien Zugang zu wissenschaftlichen Daten fest und fördern dadurch auch die freie Verfügbarkeit der Forschungsdaten im Internet. Dies ist wiederum ein Anreiz für die Forscher, ihre Daten zu veröffentlichen und damit im logischen Schluss auch in ein digitales Archiv zu überführen.

Ein Jahr später – im Jahr 2004 – forderte die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) einen besseren Zugang zu Forschungsdaten aus öffentlich geförderter Forschung (OECD 2004, Abs. 17). Mehrere Regierungen unterzeichneten eine Erklärung, die aus den folgenden Zielen und Grundsätzen bestehet: Offenheit, Transparenz, legale Konformität, Haftbarkeit, Professionalität, Wahrung des geistigen Eigentums, Interoperiabilität, Qualität & Sicherheit, Effizienz und Verantwortung (vgl. OECD 2004, Annex 1). Im Dezember 2006 folgte durch den Rat der OECD die „Empfehlung betreffend den Zugang zu Forschungsdaten aus öffentlicher Förderung“ (OECD 2006). Diese Erklärung baut zu Teilen auf den Zielen und Grundsätzen aus der Erklärung von 2004 auf und ist für die Mitgliedstaaten der OECD – also auch Deutschland – bindend und wird in naher Zukunft in die jeweils nationale Gesetzgebung implementiert. Im Abschnitt M geht es um die Nachhaltigkeit von Forschungsdaten. Die OECD schlägt hier vor: „Research funding agencies and research institutions, therefore, should consider the long-term preservation of data at the outset of each new project, and in particular, determine the most appropriate archival facilities for the data.“ (OECD 2006, Abs. M) Es bleibt abzuwarten, ob die Projekte dieser Empfehlung Folge leisten und sich zu Beginn des Forschungsprojektes um die Langzeitarchivierung der anfallenden Forschungsdaten Gedanken machen.

Ein weiteres Papier, das im letzten Jahr veröffentlicht wurde, ist die „Schwerpunktinitiative ‚Digitale Information’ der Allianz-Partnerorganisationen“ (Al-

lianz-Initiative Digitale Information 2008). Im Juni 2008 unterzeichneten die Mitglieder[23] der Allianz diese Empfehlung, um ihre Aktivitäten und Ressourcen in unterschiedlichen Bereichen zu koordinieren und zu bündeln. Ziel soll es sein, Wissenschaftlern den entgelt- und barrierefreien Zugang zu Publikationen, Forschungsdaten und virtuellen Forschungs- und Kommunikationsumgebungen zu ermöglichen. In ihrem Aktionsfeld 4 geht die Allianz detailliert auf die Problematik der Forschungsdaten ein. Es wird das Ziel verfolgt, eine Infrastruktur zum sammeln, archivieren und dem anschließenden Verfügbarmachen der Forschungsdaten den Fachdisziplinen anzubieten, die noch auf keine solche Struktur zurückgreifen können. Dadurch wird den Anforderungen der guten wissenschaftlichen Praxis entsprochen, aber auch die Nachnutzung der Daten durch Wissenschaftler zu einem späteren Zeitpunkt ermöglicht. Ein weiteres Ziel ist es, eine gemeinsame Forschungsdaten-Policy zu generieren und zu veröffentlichen, um den Wissenschaftlern den richtigen Umgang mit Forschungsdaten nahezulegen. Außerdem soll der Kontakt unter den Wissenschaftlern als Produzenten der Daten und den Informationswissenschaftlern als Datenkuratoren gestärkt werden, sowie Angebote zur Förderung von Pilotprojekten gemacht werden. Im Anschluss soll der Aufbau von institutionellen und fachspezifischen Repositorien für Forschungsdaten gefördert werden.

Die zurzeit aktuellsten Empfehlungen sind die „Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten“ der DFG (DFG 2009), die im Januar 2009 veröffentlicht wurden. Dieses Papier baut auf den Empfehlungen aus dem Jahr 1998 auf, formuliert sie aber noch deutlicher auf dem Gebiet der Forschungsdaten aus. In der Empfehlung 2 wird von einem fachspezifischen Organisationskonzept gesprochen, das die nachhaltige Speicherung der Forschungsdaten regeln soll (vgl. DFG 2009, Empfehlung 2, S. 3). Zudem sichert die DFG im Rahmen ihrer Fördermöglichkeiten finanzielle Mittel zu, damit der Ausbau von Speicher- und Nachweisstrukturen für Forschungsdaten weiter vorangebracht werden kann. Die DFG spricht sich in diesem Papier dazu aus, Explorations-Projekte durchzuführen, um weitere Erfahrungen in den verschiedenen Fachgebieten zu erhalten.

Bis jetzt bleiben erste Erfolge auf Grund dieser Empfehlungen und Policies leider aus. Wie bereits angesprochen, gibt es keinerlei Rückmeldung, inwieweit die zehn Jahre der „guten wissenschaftlichen Praxis“ bei einzelnen Institutionen berücksichtigt wurden. Dies liegt womöglich auch daran, dass es sich bei den vorgestellten Papieren um Empfehlungen handelt und nicht um

[23] Alexander von Humboldt-Stiftung, Deutscher Akademischer Austauschdienst, Deutsche Forschungsgemeinschaft, Fraunhofer-Gesellschaft, Helmholtz-Gemeinschaft Deutscher Forschungszentren, Hochschulrektorenkonferenz, Leibniz-Gemeinschaft, Max-Planck-Gesellschaft, Wissenschaftsrat

verbindliche Richtlinien. Es bleibt jedoch abzuwarten, wie sich die Umsetzung der OECD-Empfehlung in nationales Recht in den einzelnen Mitgliedstaaten vollzieht.

3 Notwendige Aspekte für die Entwicklung eines Geschäftsmodells für die Langzeitarchivierung

3.1 Beteiligte Akteure und deren Rolle im Prozess der Langzeitarchivierung

In der Literatur gibt es verschiedene Ansätze, den Begriff „Geschäftsmodelle" zu definieren. Die Autoren Scheer, Deelmann und Loos (Scheer/Deelmann/Loos 2003) stellen in ihrer Arbeit verschiedene Definitionen von Geschäftsmodellen vor. Sie kommen zu dem Schluss, dass „im Gegensatz zum Konsens über den Nutzen von Geschäftsmodellen und ihrer Anwendung [...] in der Literatur keine Einigkeit über das Wesen eines Geschäftsmodells, seiner Bestandteile und seiner Abbildung [besteht]." (Scheer/Deelmann/Loos 2003, S. 7) Sie stellen jedoch nach einer Betrachtung von verschiedenen Definitionen fest, dass die beteiligten Akteure und deren Rolle im Geschäftsverlauf in allen Ansätzen die wichtigsten Bestandteile des Geschäftsmodellbegriffs darstellen (vgl. Scheer/Deelmann/Loos 2003, S. 21f).

Aus diesem Grund werden im Folgenden die zentralen Akteure im Zusammenhang mit der Langzeitarchivierung von Forschungsdaten vorgestellt und deren Rolle im Prozess der Archivierung von digitalen Objekten dargestellt. Am Beispiel des Göttinger Wissenschaftsstandortes wird konkret eine Rollenzuweisung der Akteure vorgenommen.

Während des gesamten Lebenszyklus der Forschungsdaten kommen verschiedene Institutionen und Personen mit den digitalen Objekten in Berührung. Bei den Akteuren handelt es sich folglich um (vgl. PARSE.Insight 2009, S. 7):

- Produzenten von Forschungsdaten: Wissenschaftler
- Nutzer von Forschungsdaten: Wissenschaftler
- Anbieter von Forschungsdaten: Datenarchive, kommerzielle Anbieter, traditionelle Gedächtnisorganisationen
- Politische Entscheidungsträger und Förderer
- Forschungsförderer, Stiftungen
- Entwickler von Langzeitarchivierungssystemen und Infrastruktur
- Verleger von Forschungsdaten und Online Datenbanken

Diese können zu folgenden vier Communities zusammengefasst werden:

- Wissenschaftler
- Anbieter von Langzeitarchivierungslösungen
- Politische Entscheidungsträger und Förderer
- Verleger

Diese Communities haben nach Liz Lyon (vgl. Lyon 2007, S. 9) jeweils eigene Aufgaben, Rechte und Verantwortlichkeiten im Umgang mit den digitalen Objekten, die in Geschäftsmodellen zur Langzeitarchivierung von Forschungsdaten berücksichtigt werden müssen. Diese traditionelle Aufteilung der Aufgaben unterliegt zurzeit Veränderungen, die auf den stetig steigenden Zuwachs digitaler Forschungsdaten, auf neue Publikationsmodelle, wie z.B. Open Access, und die zunehmende Verlinkung von Forschungsdaten und Publikationen zurückzuführen sind. Diese Herausforderungen können nur im Rahmen von Kooperationen gemeistert werden. Geschäftsmodelle zur Langzeitarchivierung von Forschungsdaten müssen diesem Anspruch gerecht werden.

Am Beispiel des Göttinger Standortes werden nun Akteure vorgestellt und ihre Rollen in einer möglichen Langzeitarchivierung von Forschungsdaten dargestellt. Für eine Umsetzung der Aufgaben einer Langzeitarchivierung von Forschungsdaten am Standort Göttingen müssen die dort vorhandenen und etablierten Einrichtungen Aufgaben in dem Workflow der Langzeitarchivierung übernehmen. Dabei sollten die bereits vorhandenen Kooperationen und Kompetenzen berücksichtigt werden. Zu den Akteuren zählen die Georg-August-Universität Göttingen, die Förderer und Geldgeber, die Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG), der Geschäftsbereich Informationstechnologie (G3-7 IT) der Universitätsmedizin der Universität Göttingen, die Niedersächsische Staats- und Universitätsbibliothek (SUB) Göttingen, das Forschungsbüro und das EU-Büro, der an der Universität zuständige Beauftragte für Datenschutz, die Produzenten der Forschungsdaten und letztlich die Endnutzer der Daten.

Die **Georg-August-Universität Göttingen** zählt seit 2007 zu den bundesweit neun Exzellenzuniversitäten, denen das Potenzial für internationale Sichtbarkeit und ein Rang unter den Spitzenuniversitäten bescheinigt wird. Um diesem Ruf auch gerecht zu werden, wäre es wünschenswert, wenn die Universität Göttingen auch im Bereich der Langzeitarchivierung von Forschungsdaten ein klares Zeichen setzen würde und hier Forschungsprojekte in diesem Bereich etabliert, um in Zukunft den Wissenschaftlern eine sichere Langzeitarchivierung ihrer Forschungsdaten anzubieten.

Die Universität Göttingen ist in verschiedenen Gebieten der Langzeitarchivierung von Forschungsdaten mit einzubeziehen. Allen voran ist sie als Dienstherr der Göttinger Wissenschaftler verantwortlich für die Archivierung der dort produzierten Forschungsdaten. Zudem ist sie für die Bereitstellung von Arbeitsplätzen für die Mitarbeiter der Langzeitarchivierung verantwortlich und spielt eine wichtige Rolle bei der Bereitstellung von finanziellen Mitteln für die Aufgabe der Langzeitarchivierung von Forschungsdaten.

Forschung wird aus verschiedenen **Förderern und Geldgebern** finanziert. Zum einem sind die Unterhaltsträger der Forschenden für die notwendige

Finanzierung der Forschung zuständig, wie z.B. die entsprechende Universität. Zum anderen spielt die Finanzierung durch Drittmittel in der Forschung eine entscheidende Rolle. Hier treten Forschungsförderer wie die DFG, das BMBF oder auch die EU ein. Nicht selten sind Mischfinanzierungen des Unterhaltsträgers und von Drittmittelgebern üblich.

Die Förderer und Geldgeber haben im Rahmen der Langzeitarchivierung von Forschungsdaten zwei Aufgaben zu erfüllen. Zum einem haben sie in ihrer Rolle als Geldgeber die Möglichkeit, gewisse Vorschriften und Richtlinien zu bestimmen, wie beispielsweise die DFG in der Richtlinie zur „Sicherung guter wissenschaftlicher Praxis" (DFG 1998, siehe dazu Kapitel 2.3). Zum anderen beteiligen sich die Förderer an der finanziellen Unterstützung der Langzeitarchivierung. Die Finanzierung kann als Anschub- oder Betriebsförderung erfolgen (dazu mehr in Kapitel 3.2.3).

Die **Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen** (GWDG) ist eine gemeinsame Einrichtung der Georg-August-Universität Göttingen und der Max-Planck-Gesellschaft. Die GWDG fungiert als Rechenzentrum für beide Einrichtungen. Zudem beteiligt sie sich an verschiedenen Forschungsaufgaben im Bereich der Angewandten Informatik.

Die GWDG spielt in der Langzeitarchivierung von Forschungsdaten die Rolle des Betreibers des physikalischen digitalen Langzeitarchivs. Bei ihr werden die digitalen Objekte abgeliefert und die Daten auf großen Speichersystemen vorgehalten.

Eine der wichtigsten Aufgaben der GWDG ist die Gewährleistung der Bitstream-Preservation der digitalen Objekte, die als Grundvoraussetzung für die Langzeitarchivierung gesehen werden kann. Weitere Aufgaben des Rechenzentrums sind die Sicherstellung der technischen Benutzbarkeit der archivierten Daten, die Sicherung der Authentizität, die technische Prüfung der eingehenden Daten sowie die Erstellung von technischen Metadaten.

Der **Geschäftsbereich Informationstechnologie (G3-7 IT) der Universitätsmedizin der Georg-August-Universität Göttingen** bietet verschiedene Dienstleistungen im IT-Bereich. Im Aufgabenbereich der G3-7 IT liegt unter Anderem der Betrieb der Server der Universitätsmedizin und die Aufrechterhaltung der Netzinfrastruktur und die Sicherheit der Netze. Der Bereich G3-7 IT ist demnach Hauptansprechpartner für alle Fragen der Informationstechnologie an der Universitätsklinik.

Damit stellt der Bereich ein zweites Rechenzentrum auf dem Göttinger Campus dar, das bereits jetzt als redundanter Speicher für die GWDG und umgekehrt agiert. Bei der G3-7 IT liegen zudem Kompetenzen im Bereich der Speicherung von personenbezogenen Daten.

Die **Niedersächsische Staats- und Universitätsbibliothek (SUB) Göttingen** hat neben ihrer Funktion als Universitätsbibliothek noch weitere Funk-

tionen, wie beispielsweise im Rahmen des Sondersammelgebietsprogramms (SSG) der DFG. Die Aufgaben im Rahmen des SSG-Programms bestehen vor allem in einer umfassenden Sammlung, Erschließung und nationalen wie internationalen Bereitstellung von themenorientierten Informationen für Wissenschaft, Forschung und Lehre. In diesem Auftrag spielen sowohl gedruckte als auch elektronische Publikationen und Medien eine wichtige Rolle. Eine weitere Besonderheit der SUB Göttingen sind ihre ausgeprägten Aktivitäten im Bereich der Forschung und Entwicklung.

Die SUB konnte am Standort Göttingen schon erste praktische Erfahrungen in der digitalen Langzeitarchivierung sammeln. So war die SUB Projektpartner beim kopal Projekt (siehe Kapitel 3.3.1). Gemeinsam mit der Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen betreibt sie das kopal-System. Weiterhin ist die SUB Göttingen in zahlreichen nationalen und internationalen Forschungsprojekten zum Thema Langzeitarchivierung beteiligt.[24]

Die Bibliothek übernimmt bei der Langzeitarchivierung von Forschungsdaten größtenteils fachbezogene Beratungstätigkeiten und Schulungen sowie Aufgaben bei der Konzeptionierung und Umsetzung organisatorischer Prozesse. Zudem kann die SUB die Entwicklung von Workflows, LZA-Leitbildern und fachbezogenen Policies zur digitalen Langzeitarchivierung von Forschungsdaten aktiv unterstützen. In der fachlichen Beratungstätigkeit kann die SUB auf ihre Kompetenzen im Bereich der Fachreferate zurückgreifen und die Produzenten der Forschungsdaten bei der Generierung der digitalen Objekte und dem richtigen Umgang mit ihnen unterstützen. Durch den direkten und unmittelbaren Kontakt der Fachreferenten zu Wissenschaftlern und auch Studierenden aus ihrem Fachbereich sind die Fachreferenten beispielsweise für die Aufgabe des Community Watch sehr gut qualifiziert. Zudem zählen auch die Entwicklung von Tools zur Verbesserung und Unterstützung von Migrationen, die Umsetzung von Migrationsprozessen, die Entwicklung von Emulatoren sowie der Technology Watch zum Aufgabenbereich der SUB (bereits im Zusammenhang mit kopal).

Durch die bereits vorhandenen Kompetenzen im Bereich der Nutzung und dem Zugang zu wissenschaftlichen Daten kann die SUB die Forschungsdaten, ähnlich wie die TIB Hannover (siehe Kapitel 2), in ihrem Katalog anbieten und so den Nutzern zur Verfügung stellen. Bei Fragen der Endnutzer zur Auffindbarkeit und Benutzung der Forschungsdaten hilft das geschulte Personal vor Ort mit Support.

---

[24] Siehe dazu auch die Webseite der Abteilung F&E der SUB Göttingen: http://rdd.sub.uni-goettingen.de/

Das Einwerben von Drittmitteln zur Forschungsförderung ist die Grundlage für eine Großzahl von Forschungsprojekten und wird vom **Forschungsbüro und EU-Büro** übernommen. Zur effektiven Einwerbung von Mitteln bedarf es vielfach einer Beratung über die verschiedenen Fördermöglichkeiten und zu Fragen der Antragsstellung.

Am Wissenschaftsstandort Göttingen ist neben dem Forschungsbüro auch ein EU-Büro für Fragen und Beratung zur Forschungsförderung der Wissenschaftler verfügbar. Diese Einrichtungen bieten Unterstützung bei der Erwerbung von Drittmitteln (z.B. bei der EU, der DFG, dem BMBF und bei Stiftungen) entweder für Forschungsprojekte, die sich mit der Fragestellung der Langzeitarchivierung von Forschungsdaten beschäftigen, oder auch bei Forschungsprojekten, bei denen Forschungsdaten anfallen, so dass finanzielle Mittel für die dauerhafte Verfügbarkeit mit beantragt werden.

Sobald personenbezogene Daten verarbeitet werden, benötigt diese Einrichtung einen **Datenschutzbeauftragten**. Die wesentlichen Aufgaben des Datenschutzbeauftragten sind im Bundesdatenschutzgesetz und im niedersächsischen Datenschutzgesetz geregelt. Er ist insbesondere für die Einhaltung dieser Gesetze und anderer Vorschriften zum Datenschutz zuständig und überwacht die ordnungsgemäße Anwendung der eingesetzten IT-Verfahren. Bei Anfragen zum Thema Datenschutz erteilt er Auskünfte und ist für die ordnungsgemäße Einführung in Bezug zum Datenschutz bei neuen IT-Verfahren zuständig.

Auch die Universität Göttingen hat einen Datenschutzbeauftragten. Dieser sollte Aufgaben in rechtlichen Fragen der Langzeitarchivierung von Forschungsdaten übernehmen. Dabei ist beispielsweise an Fragen der rechtlichen Klärung bei medizinischen Forschungsdaten zu denken, die durch ihren möglichen Personenbezug nicht ohne Weiteres gespeichert werden dürfen.

Forschungsdaten entstehen in den verschiedensten Konstellationen an einer Hochschule. Zunächst ist an den einzelnen Forscher als **Produzenten von Forschungsdaten** zu denken. In den meisten Fällen wird jedoch in Forschungsprojekten gearbeitet, die an einzelnen Institutionen oder auch institutsübergreifend organisiert sein können. Ein Beispiel für institutsübergreifende Projekte sind die Sonderforschungsbereiche der DFG. Zudem können Forschungsdaten auch im Rahmen von nationalen oder internationalen Forschungskooperationen entstehen.

Die Göttinger Wissenschaftler in Forschungsprojekten, die Professoren und die Mitarbeiter an ihren Lehrstühlen sind somit Produzenten der Forschungsdaten, die später in ein Langzeitarchiv überführt werden sollen. Die Forschenden müssen deshalb bereits bei der Beantragung von Forschungsmitteln berücksichtigen, dass bei ihren Arbeiten Forschungsdaten anfallen werden, die später in eine dauerhafte Archivierung überführt werden sollen. Aus diesem Grund sollte zu Beginn eines Forschungsvorhabens eine Abschät-

zung der ungefähren Menge und der relevanten Daten-Formate geschehen. Diese Menge kann je nach Fachrichtung sehr unterschiedlich ausfallen. Die Produzenten sollten vor der Abgabe ihrer Forschungsdaten die Aufgabe erfüllen, die digitalen Objekte mit deskriptiven Metadaten anzureichern, da dieser Arbeitsschritt am einfachsten von den Produzenten der Daten selbst erledigt werden kann und eine spätere Generierung der Metadaten äußerst kostspielig ist.

Die möglichen **Endnutzer** der langzeitarchivierten Forschungsdaten überschneiden sich größtenteils mit den zuvor genannten Produzenten der Daten. Dies sind die Wissenschaftler, die beispielsweise anhand bereits vorliegender Daten weiterforschen möchten. Denkbar sind auch Nutzer aus der Öffentlichkeit, die Zugang zu den Forschungsdaten suchen, vermutlich aber in geringerem Umfang als Wissenschaftler. Ein Beispiel wäre ein Patient aus einem medizinischen Forschungsprojekt, der Zugang zu seiner digitalen Patientenakte haben möchte.

Die Gruppe der Endnutzer hat die Aufgabe, das System der Langzeitarchivierung durch Feedback ständig zu verbessern und durch neue Anforderungen an die Entwickler zur Optimierung beizutragen.

Der **Göttinger Universitätsverlag** wurde im Sommer 2003 gegründet, um primär den Göttinger Wissenschaftlern Publikationsleistungen anzubieten. Der Verlag bietet den größten Teil seiner Publikationen digital im Internet als Open Access an und alle Veröffentlichungen sind über den Katalog der SUB Göttingen recherchierbar. Die im Verlag veröffentlichten Publikationen werden in das kopal-System überführt und sind damit an die Langzeitarchivierungsinfrastruktur angebunden.

Im Bereich der Publikation von Forschungsdaten sind im Göttinger Universitätsverlag noch keine Unternehmungen angestellt wurden.

Zusammenfassung

Es wird für den Standort Göttingen von großer Bedeutung sein, die bereits umfangreichen Kompetenzen bei den verschiedenen Einrichtungen zu nutzen. So sollte jede Institution ihre Erfahrungen und auch Kontakte nutzen, um die Langzeitarchivierung von Forschungsdaten voranzutreiben. Eine Kooperation von allen Beteiligten wird in Göttingen eine Notwendigkeit sein, um die umfangreichen Aufgaben, die die Langzeitarchivierung mit sich bringt, zu meistern. Um einige dieser Aufgaben, die für ein Geschäftsmodell notwendig sind, wird es im nächsten Kapitel gehen.

3.2 Komponenten für den Aufbau eines Geschäftsmodells für die Langzeitarchivierung

Damit ein Geschäftsmodell für die Langzeitarchivierung von Forschungsdaten umgesetzt werden kann, sind mehrere Komponenten notwendig. Zum einen ist dies gesamte technische Infrastruktur, die für den Betrieb des Ar-

chivs notwendig ist. Dazu stehen die verschiedenen LZA-Systeme zur Auswahl und müssen durch den einzelnen Dienstleister nach den speziellen Anforderungen am Standort ausgewählt werden. Eine Auflistung von verschiedenen technischen LZA-Systemen findet sich bei Borghoff (Borghoff 2005). Zum anderen sind umfangreiche organisatorische und administrative Komponenten für eine Langzeitarchivierung notwendig.

In diesem Abschnitt werden die entstehenden Kosten bei der Langzeitarchivierung von Forschungsdaten genauer betrachtet und eine mögliche Berechnungsformel vorgestellt. Diese Kosten können bei den unterschiedlichen Arbeitsschritten innerhalb sowie außerhalb des LZA-Systems anfallen. Anschließend werden Service-Level-Agreements betrachtet, die als Ausformulierungen von Verträgen zwischen den LZA-Dienstleister und dem Datenproduzenten verstanden werden können. Sie dienen beiden Vertragsseiten dazu, die Dienstleistung transparent zu gestalten. Abschließend werden mögliche Finanzierungsmodelle für eine Langzeitarchivierung von Forschungsdaten sowie zwei unterschiedliche Finanzierungsarten dargestellt.

3.2.1 Kostenmodelle

Die anfallenden Kosten bei der Langzeitarchivierung von digitalen Objekten sind ein wichtiger Punkt, da ohne eine Kostenabschätzung des Vorhabens „Langzeitarchivierung“ keine Finanzierung erfolgen wird, ohne die die besten Lösungen zur Langzeitarchivierung nicht umgesetzt werden können. So stellt Ashley sehr passend fest: „However, the main barrier that many face is the difficulty of deciding how much such a service might cost.“ (Ashley 1999)

Der Aspekt der auftretenden Kosten ist umso gravierender, als Forscher diesen Aspekt größtenteils aus ihrer wissenschaftlichen Arbeit ausklammern: Nach einer Umfrage[25] antworteten nur 20% der Befragten, sich bereits bei der Konzeption ihres Projektes mit dem Thema „Kosten der Langzeitarchivierung“ beschäftigt zu haben (vgl. Abbildung 6). Auch dieses Ergebnis der Befragung zeigt, dass die Wichtigkeit der anfallenden Kosten im Rahmen der Langzeitarchivierung unterschätzt wird. Dies betrifft nicht in erster Linie den Forscher selbst, sondern viel mehr die Förderer und die Politik. An den Wissenschaftsstandorten muss eine Ansprechperson vorhanden sein, die über die anfallenden Kosten bei der Langzeitarchivierung von Forschungsdaten umfassend informiert ist und für mehr Verständnis bei den Produzenten der Daten werben kann, um allmählich ein Paradigmenwechsel in der Wissenschaftskultur zu erzielen.

[25] Des Münchener Projektes zur Langzeitarchivierung von Digitalisaten (Lang 2008).

Abbildung 6: Welche Anforderungen der Langzeitarchivierung haben Sie bereits bei der Konzeption des Projektes berücksichtigt? (Quelle: Lang 2008, Folie 14)

Kosten fallen im Prozess der Langzeitarchivierung in unterschiedlichen Arbeitsschritten an und hängen dabei von verschiedenen Merkmalen ab. Mit dem Thema haben sich bereits einige Projekte und Einrichtungen auseinandergesetzt.

Die Kosten, die durch eine vertrauenswürdige Langzeitarchivierung entstehen, lassen sich nach Wollenschläger (vgl. Wollenschläger 2008, 5-1f.) in drei unterschiedliche Arten gliedern:

- *Initiale Kosten*: Entstehen bei der Informationsbeschaffung und bei der Projektplanung zu einem LZA-System
- *Beschaffungskosten*: Entstehen bei Beschaffung von Hard- und Software, aber auch bei der Neueinstellung von Personal oder beim Bau von neuen Räumlichkeiten
- *Betriebskosten*: Entstehen beim dauerhaften Betrieb des LZA-Systems

Mehrere Merkmale sind für die konkreten Kosten verantwortlich:

- die Menge der digitalen Daten
- die Heterogenität der vorliegenden Dateiformate
- die Qualität der Metadaten
- die geforderte Zugriffsmöglichkeit und die damit verbundenen Schnittstellen zu den digitalen Objekten
- die rechtlichen Verpflichtungen, zur Archivierung der Daten

Eine detaillierte Berechung von Kosten zur Langzeitarchivierung findet sich bei dem Projekt Life Cycle Information for E-Literature[26] (LIFE), das sich mit dem Lebenszyklus von digitalen Objekten beschäftigt. Das Projekt ist vom Joint Information Systems Committee (JISC) gefördert. Beteiligte Projektpartner sind das University College London Library Services und die British Library. Das Projekt bestand aus zwei Projektphasen: In der ersten Phase (LIFE^{1}) wurden die Lebenszyklen von digitalen Objekten untersucht und jeder der einzelnen Abschnitte im Lebenszyklus mit Kosten belegt, um am Ende eine Gesamtkostenberechnung für die Langzeitarchivierung von digitalen Objekten angeben zu können. In der zweiten Phase des Projektes (LIFE^{2}) wurde die Formel zur Berechnung der Kosten anhand der Einbeziehung von drei Fallstudien verfeinert .

LIFE hat zur Berechnung der Kosten in seiner ersten Projektphase eine Formel aufgestellt, die innerhalb der zweiten Projektphase einer leichten Modifikation unterzogen wurde (vgl. Ayris et al. 2008, S. 18). Dabei lassen sich die Kosten wie folgt berechnen:

$$C + Aq_T + I_T + BP_T + CP_T + Ac_T = L_T$$

C (Creation or Purchase): Kosten bei der Erstellung oder der Erwerbung

Aq_T (Acquisition): Kosten für die Erfassung der Daten für das Archiv im Zeitraum T

I_T (Ingest): Kosten für die Aufnahme der Daten in das LZA-System im Zeitraum T

BP_T (Bit-stream Preservation): Kosten für die Bitstream-Preservation (technisch) im Zeitraum T

CP_T (Content Preservation): Kosten für die inhaltliche Langzeiterhaltung der Daten im Zeitraum T

Ac_T (Access): Kosten für den Zugriff auf das digitale Objekt

L_T (Lifecycle Cost): gesamte Lebenszykluskosten eines digitalen Objektes im Zeitraum T

Abbildung 7: Kostenformel nach LIFE^{2} (Quelle: Ayris 2008, S. 18)

Bei LIFE sind zu den einzelnen Lebenszyklus-Stufen (Lifecycle Stages) noch Elemente (Lifecycle Elements) benannt, um die einzelnen Stufen noch detaillierter zu beschreiben (vgl. Tabelle1). So lässt sich beispielsweise der Arbeitsschritt „Aufnahme in das Archiv“ (Ingest) in fünf Elemente aufteilen:

- Qualitätskontrolle (Quality Assurance)
- Anreichern der digitalen Objekte mit Metadaten (Metadata)
- Ablegen der Daten in das Archiv (Deposit)
- Bestandsaktualisierung (Holdings Update)
- Nachweisverbindung herstellen (Reference Linking)

---

[26] http://www.life.ac.uk/

Diese Elemente werden wiederum mit Unterelementen (Lifecycle Subelement) näher beschrieben. Das verdeutlicht das Beispiel des Elements Metadaten:

- Datei-Format identifizieren (File Format Identification)
- Datei-Format validieren und auf Integrität prüfen (File Format Validation and Integrity Check)
- Metadaten automatisch extrahieren und speichern (Metadata Extraction and Recording)
- Manuelle Erstellung der Metadaten (Metadata Creation)
- Event Metadaten ergänzen (Record Event Metadata)

Durch diese detaillierte Aufstellung der einzelnen Stufen in Elemente und in Unterelemente wird es möglich, eine genaue Untergliederung der einzelnen Prozesse zu erhalten, denen im Anschluss Kosten zugewiesen werden können.

Lifecycle Stage	Creation or Purchase		Acquisition	Ingest	Bit-stream Preservation	Content Preservation	Access
Lifecycle Elements	...		Selection	Quality Assurance	Repository Administration	Preservation Watch	Access Provision
	...		Submission Agreement	Metadata	Storage Provision	Preservation Planning	Access Control
	...		IPR & Licensing	Deposit	Refreshment	Preservation Action	User Support
	...		Ordering & Invoicing	Holdings Update	Backup	Re-Ingest	
	...		Obtaining	Reference Linking	Inspection	Disposal	
	...		Check-In				

Tabelle 1: Das LIFE-Modell V2 (Quelle: Ayris et al. 2008)

Zusätzlich zu den Lebenszykluskosten, die durch das LIFE Projekt für das digitale Objekt identifiziert wurden, fallen bei dem LZA-Dienstleister noch Betriebskosten an, die nicht direkt den einzelnen Objekten zugerechnet werden können. Innerhalb des LIFE Projekts spricht man bei diesen Kosten von Non-Lifecycle Costs, die sich in die folgenden drei Stufen einteilen lassen:

- Management und Administration (Management and Administration)
- System und Infrastruktur (System/Infrastructure)
- Ökonomische Anpassungen (Economic Adjustments)

In der JISC-geförderten Studie „Keeping Research Data Save“ beschäftigen sich die Autoren Beagrie, Chruszcz und Lavoie (Beagrie/Chruszcz/Lavoie

2008) mit dem Thema der Kosten bei der Langzeitarchivierung von Forschungsdaten am Beispiel von Forschungseinrichtungen aus Großbritannien.

In der Studie wird eine Verteilung der Kosten wie folgt dargestellt: 42% der entstehenden Kosten für eine Langzeitarchivierung fallen auf die Stufen Erfassung und Aufnahme der Daten, 23% fallen auf die Stufen Bitstream-Preservation und inhaltliche Langzeiterhaltung an und 35% der Kosten entstehen beim Zugriff (vgl. Abbildung 8).

Abbildung 8: Kostenarten nach Beargrie, Chruszcz und Lavoie (Beagrie 2008, S. 72)

Ashley (Ashley 1999) beschreibt zudem, dass der größte Teil der Kosten – ca. 70% – bei den Personalkosten anfallen werden.

Ein weiterer zentraler Aspekt bei der Kostenentwicklung wird die Anzahl der Datei-Formate sein, die vom LZA-System unterstützt werden sollen bzw. müssen. Dabei spielen auch wieder die Personalkosten eine wichtige Rolle, da qualifizierte Mitarbeiter für verschiedene Datei-Formate erforderlich sind, die technisch und fachlich mit dem Format vertraut sind. Vermutlich wird es für den LZA-Dienstleister sinnvoll sein, nicht jedes Datei-Format von dem Auftraggeber anzunehmen, sondern auf allgemein standardisierten Formaten zu bestehen; besonders unter dem Gesichtspunkt, dass es sich bei Forschungsdaten um sehr heterogene Daten handelt. Dies zeigt auch die Studie von Björk (Björk 2007), die im Rahmen des LIFE Projektes stattfand. Danach haben die Datei-Formate PDF, TXT und HTML ca. 85% aller Objekte abgedeckt, waren damit jedoch für nur 7% der Kosten verantwortlich. Währendessen wurden 41% der Kosten durch die 12 am wenigsten verwendeten Formate (0,1%) hervorgerufen.

Ein möglicher Aspekt für Kosteneinsparungen wird der Zeitpunkt sein, zu dem die Metadaten dem digitalen Objekt hinzugefügt werden (vgl. DPC

2008). Bei der Erstellung der digitalen Daten ist es wesentlich einfacher, d.h. kostengünstiger, dem Datenobjekt deskriptive Metadaten mitzugeben, als zu einem späteren Zeitpunkt, wenn sich ein weiterer Fachwissenschaftler mit dem Objekt beschäftigen müsste, um die deskriptiven Metadaten zu ergänzen, ohne den Kontext des erstellten Objekts genau zu kennen.

In der Studie „Keeping Research Data Safe" (Beagrie 2008, S. 6) wird beschrieben, dass das Überarbeiten von zuvor schlechten Metadaten bei 10.000 Datensätzen nach einer Periode von 10 Jahren ungefähr 10.000 € kosten würde, im Gegensatz zu ca. 333 € bei sofortiger Beschreibung der Metadaten. Dies entspricht einem Faktor von 30.

Die vorgestellten Modelle zeigen, dass es im Bereich der Kostenabschätzung der Langzeitarchivierung bereits erste nennenswerte Ansätze gibt. Auch im speziellen Bereich der Forschungsdaten, beschäftigen sich Projekte und Studien mit diesem Thema.

3.2.2 Service-Level-Agreements

Die rechtliche Situation zwischen dem Anbieter der Dienstleistung „Langzeitarchivierung von Forschungsdaten" einerseits und dem Auftraggeber[27] als dem Produzenten der digitalen Objekte andererseits muss vertraglich abgesichert sein. Dafür schließen die beiden Parteien einen Dienstleistungsvertrag ab. Innerhalb der vertraglichen Regelung stellt ein Service-Level-Agreement (SLA) darin den beschreibenden Teil des Vertrages dar. Mit einem SLA werden der Servicegrad, die Qualität und die Rechte, Pflichten und Zielvereinbarungen zwischen dem Dienstleister und dem Auftraggeber festgelegt. Auf diese Weise wird ermöglicht, dass im Anschluss beide Parteien die Einhaltung dieser Zielvorgaben kontrollieren können. Sollten diese Ziele nicht erreicht werden, treten zuvor festgelegte Sanktionen in Kraft.

Ein Vorteil von solchen Service-Level-Agreements ist die Transparenz der angebotenen Dienstleistung für den Auftraggeber. Diese Transparenz ist wichtig, um die Vertrauenswürdigkeit des LZA-Systems zu gewährleisten. So sieht der Auftraggeber, wie mit seinen Daten im Folgenden umgegangen wird (vgl. nestor-Arbeitsgruppe Vertrauenswürdige Archive – Zertifizierung 2008, S. 8) und für welche Leistungen welche Kosten anfallen. Zudem kann der Auftraggeber sich seine Dienstleistungen individuell nach seinen Anforderungen an die Forschungsdaten, die zu archivieren sind, zusammenstellen und muss nicht für Leistungen zahlen, die er gar nicht in Anspruch nehmen will oder muss. Durch die genaue Formulierung der Leistungen werden Missverständ-

---

[27] Der Auftrag kann die unterschiedlichen Konstellationen haben. So ist denkbar, dass ein Forschungsprojekt als Auftraggeber an den LZA-Dienstleister herantritt oder aber ein gesamter Wissenschaftsstandort, der einen Vertrag schließt, der alle am Standort forschenden Wissenschaftler beinhaltet.

nisse im Vorhinein verhindert. Zudem besteht eine Planungssicherheit für den Dienstleister der Langzeitarchivierung, da ihm seine Aufgaben im Vorfeld bewusst sind.

SLA's umfassen:

- Leistungen, die exakt formuliert werden und messbar sind
- eine detaillierte Leistungsbeschreibung
- eine Bestimmung des mit dem SLA verfolgten Zwecks
- die Verantwortung der Dienstleister
- die Verantwortung der Auftraggeber
- die geforderte Verfügbarkeit der Dienstleistung
- die Sanktionen bei Nichteinhaltung
- die Preisgestaltung

LZA-Dienstleister

Service-Level-Agreement
- Leistungsbeschreibung
- Ziele
- Rechte & Pflichten
- Verfügbarkeit
- Sanktionen
- Preisgestaltung

Auftraggeber

Abbildung 9: Beziehungen bei Service-Level-Agreements

Der LZA-Dienstleister hat durch die Verwendung von SLAs die Möglichkeit, seinem Auftraggeber eine transparente und individuelle Dienstleistung anzubieten. Die einzelnen Service-Level können dabei sehr allgemeine Leistungen beschreiben, wie beispielsweise der Art der technischen Aufbewahrung der digitalen Objekte in Form einer „einfachen" Bitstream-Preservation, sowie auch komplexere Vorgänge, wie z.B. einer betreuten Archivierung mit regelmäßigen Formatmigrationen. Der LZA-Dienstleister hat darüber hinaus durch die SL auch die Option, Leistungen sehr genau zu beschreiben, wie beispielsweise die Betreuung der Datenproduzenten durch den Anbieter. Der Support kann entweder per E-Mail angeboten werden, durch eine Telefon-Hotline oder durch einen Ansprechpartner vor Ort. Durch diese Möglichkeiten hat der LZA-Anbieter ein Portfolio, dass er auf die Wünsche des Auftraggebers angepasst optimal anbieten kann.

Ein weiterer wichtiger Aspekt bei der Ausgestaltung von Service-Level ist die Frage, ob ein bestehender Vertrag mit der Zeit Anpassung erhalten kann. Bei dem Auftraggeber können Veränderungen in Strukturen neue Anforderungen an die Langzeitarchivierung seiner Daten erfordern. Denkbar sind längere Vorhaltezeiträume, oder der Auftraggeber entscheidet sich auf Grund von

ersten Erfahrungen mit den langzeitarchivierten Daten, doch regelmäßige Formatmigrationen durchführen zu lassen, die im vorherigen Vertrag nicht vorgesehen waren. Alle diese Beispiele und noch viele weitere sind sehr wahrscheinlich, da sich die Langzeitarchivierung von Forschungsdaten schon aus der Sache heraus über einen langen Zeitraum erstreckt. Diese Aspekte müssen bei der Gestaltung von SL mit berücksichtigt werden.

Das Online Computer Library Center (OCLC) beschäftigt sich schon seit längerem mit dem Thema der digitalen Langzeitarchivierung in verschiedensten Bereichen. Durch Teilnahme an Komitees, Weiterbildungsveranstaltungen und Beteiligungen an Forschung auf diesem Gebiet konnten Erfahrungen auf dem Gebiet des Umgangs und der Langzeitarchivierung von digitalen Objekten gesammelt werden. Seit 2002 bietet OCLC das *OCLC Digital Archive* für Bibliotheken, Archive, Museen und Bildungseinrichtungen an, da in diesen Einrichtungen eine vermehrte Nachfrage nach vertrauenswürdiger Langzeitarchivierung von digitalen Objekten aufkam (vgl. OCLC 2006).

In der Preservation Policy von OCLC (OCLC 2006) finden sich unter anderem auch Service-Level, die die Leistungen von dem von OCLC angebotenen Dienst genauer beschreiben.

So wird das Service-Level „*Bit Preservation*“ angeboten. Hierbei werden die digitalen Objekte inklusive ihrer Metadaten in das digitale Archiv überführt und dort in ihrer ursprünglichen Form gespeichert. Ein Zugriff auf und die Verwaltung von Daten ist nach der Speicherung im Archiv dem Datenlieferanten und den vorher vertraglich festgelegten Institutionen erlaubt. Beim Ingest der Daten und danach jährlich werden die digitalen Objekte auf ihre Beständigkeit und Viren überprüft.

Ein zweites Service-Level, das bereits vom *OCLC Digital Archive* angeboten wird, ist die „*Local Preservation*“. Dieses Service-Level ist mit einigen Ausnahmen identisch mit dem „*Bit Preservation*“ Level. Zusätzlich können die Daten in ein lokales Repositorium überführt werden, wo sie für jeden zugänglich sind. Zudem werden die digitalen Objekte im METS-Format im Digital Archive abgelegt.

Ein drittes Service-Level ist die „*Full Preservation*“, die jedoch noch nicht durch OCLC angeboten wird. Dieses Service-Level zielt auf eine „wirkliche“ Langzeitarchivierung des digitalen Objektes, was nach OCLC mehr als nur die technische Integrität der Daten beinhaltet (vgl. OCLC 2006, S. 5). Hierzu zählt unter anderem ein Technology Watch, um Veränderungen an Datei-Formaten zu erkennen und ggf. Maßnahmen einleiten zu können, damit das Format auch in Zukunft noch lesbar und interpretierbar ist. Eine weitere Maßnahme wäre es, einen detaillierten Preservation Plan aufzustellen, in dem die Strategie zum weiteren Umgang mit dem digitalen Objekt beschrieben wird.

Bei der Aufstellung von Service-Level geht OCLC nicht detailliert auf alle zuvor genannten Punkte für die Bestimmung von SL ein. So werden beispielsweise die Sanktionen bei Nichteinhalten der Leistungen nicht genannt, noch werden mögliche Kosten für die Leistungen erwähnt.

Im Rahmen des KoLaWiss-Projektes, das sich mit einem Geschäftsmodell für den Standort Göttingen beschäftigte, beinhaltete das Arbeitspaket „Organisation" das Thema Service-Level. Hier wurde weniger auf die technischen Aspekte wie bei dem zuvor genannten Beispiel von OCLC eingegangen, als vielmehr der zeitliche Aspekt der Langzeitarchivierung betrachtet. Die Vorhaltezeiträume der Forschungsdaten werden bei KoLaWiss in vier Service-Level unterteilt, wobei je nach Länge des Vorhaltezeitraums der Kostenaufwand für die Langzeitarchivierung steigt (vgl. Metje 2008, siehe Abbildung 10).

Das erste SL „*Backup*" sieht eine kurze Aufbewahrungsdauer der digitalen Forschungsdaten vor. Gerechnet wird mit einer Zeitspanne von fünf Jahren. Dieses SL ist für Forschungsdaten mit einer kurzfristigen Relevanz. Dies sind beispielsweise Forschungsdaten, die nur für die Dauer des aktuellen Forschungsprojektes notwendig sind.

Ein zweites Service-Level *„Absicherung guter wissenschaftlicher Praxis"* verfolgt die Einhaltung der Richtlinie der DFG zur guten wissenschaftlichen Praxis, die eine Nutzung der Daten innerhalb von zehn Jahren vorsieht. Dieses Service-Level wird auf eine Vielzahl von Forschungsdaten zutreffen, die von öffentlichen Förderorganisationen finanzielle unterstützt werden.

Bei einem weiteren SL *„Rechtssicherheit"* geht es um gesetzliche Speicherfristen, die je nach Vorlage unterschiedlich ausfallen können. Unter dieses Service-Level fallen beispielsweise anonymisierte Patientendaten für medizinische Forschungsdaten oder Röntgenbilder, die 30 Jahre aufbewahrt werden müssen.

Das letzte Service-Level *„Sicherung des wissenschaftlichen und kulturellen Erbes"* beinhaltet die zeitlich unbegrenzte Nutzbarkeit von digitalen Forschungsdaten, die für die Nachwelt als wissenschaftliches und kulturelles Erbe gelten. Bei diesen Forschungsdaten muss davon ausgegangen werden, dass sie nicht reproduzierbar sind, gleichzeitig für zukünftige wissenschaftliche Fragestellungen von hoher Bedeutung sind, oder aufgrund hoher finanzieller Aufwendungen für ihre Erstellung als unverzichtbar angesehen werden müssen.

Kostenaufwand

Service-Level:
Sicherung des wissenschaftlichen und kulturellen Erbes

Service-Level:
Rechtssicherheit

Service-Level:
Absicherung guter wissenschaftlicher Praxis

Service-Level:
„Backup"

Aufbewahrungsdauer

Abbildung 10: Service-Level nach KoLaWiss (Quelle: Metje 2008)

Service-Level müssen von dem jeweiligen LZA-Anbieter einheitlich definiert werden, damit eine Transparenz nach außen sichtbar wird, und damit für den Auftraggeber ein Vergleich der verschiedenen Service-Level möglich wird. Anhand der zu Beginn genannten Elemente eines Service-Level-Agreements könnte ein Service-Level wie folgt aussehen:

Titel des Service-Level	
Kurze Beschreibung, Ziele des SL	
Detaillierte Beschreibung der Leistungen	
Rechte des Anbieters	Rechte des Auftraggebers
Pflichten des Anbieters	Pflichten des Auftraggebers
Geforderte zeitliche Verfügbarkeit der Dienstleistung	
Sanktionen	
Preisgestaltung	

Tabelle 2: Beispiel für ein Service-Level

Neben eher allgemeinen Service-Level, wie zuvor bei OCLC und KoLaWiss beschrieben, werden noch weitere SL notwendig sein. Diese beschreiben und bewerten die einzelnen Tätigkeiten, die im Prozess der Langzeitarchivierung von Forschungsdaten auftreten. Diese granulare Aufteilung der einzelnen Prozesse kann zum Beispiel an den im vorherigen Kapitel genannten Lebenszyklus-Stufen und -Elementen vom LIFE-Projekt erfolgen. So können zu jedem Element Service-Level erarbeitet werden, die dieses Element mit verschiedenen SL beschreiben. Eine weitere denkbare Herangehensweise wäre es, die Service-Level anhand des OAIS-Referenzmodells näher zu gliedern und dort beispielsweise zu den einzelnen Punkten SL zu beschreiben.

Exemplarisch könnte ein Service-Level für das Erstellen von Metadaten wie folgt aussehen:

<table>
<tr><th colspan="2">Metadaten I</th></tr>
<tr><td colspan="2">Die technischen und deskriptiven Metadaten zum Forschungsdatensatz werden von dem Produzenten der Daten selbst erstellt.</td></tr>
<tr><td colspan="2">Alle Metadaten – sowohl technische als auch deskriptive Metadaten – werden von dem Produzenten der Daten selbst dem Datensatz beigefügt. Dazu kann das bereitgestellte Tool verwendet werden.</td></tr>
<tr><td>Der Anbieter muss die Daten nicht auf die Richtigkeit hin überprüfen.</td><td></td></tr>
<tr><td>Der Anbieter fügt die LZA-spezifischen Metadaten zum Datensatz hinzu.</td><td>Der Auftraggeber muss sich bei der Eingabe der Metadaten an die Vorgaben aus dem Handbuch halten.</td></tr>
<tr><td colspan="2">Die Metadaten müssen zusammen mit dem digitalen Datensatz abgegeben werden.</td></tr>
<tr><td colspan="2">Werden durch den Produzenten nicht die vorgeschriebenen Metadaten mitgeliefert, dann wird der Datensatz nicht in das LZA-System überführt. Er hat die Möglichkeit eine Korrektur vorzunehmen.</td></tr>
<tr><td colspan="2">Dieses Service-Level ist am kostengünstigsten.</td></tr>
</table>

Tabelle 3: Service-Level: Metadaten I

<table>
<tr><th colspan="2">Metadaten II</th></tr>
<tr><td colspan="2">Die deskriptiven Metadaten zum Forschungsdatensatz werden von dem Produzenten der Daten selbst erstellt.</td></tr>
<tr><td colspan="2">Die deskriptiven Metadaten werden vom Produzenten der Daten selbst dem Datensatz beigefügt. Dazu kann das bereitgestellte Tool verwendet werden. Zudem werden die deskriptiven Metadaten seitens des Anbieters stichpunktartig auf Richtigkeit überprüft.</td></tr>
<tr><td>Der Anbieter muss die Daten nicht auf</td><td>Der Auftraggeber hat das Recht, die</td></tr>
</table>

<table>
<tr><th colspan="2">Metadaten II</th></tr>
<tr><td>die Richtigkeit hin überprüfen (lediglich stichpunktartige Überprüfungen).</td><td>technischen Metadaten auf Richtigkeit zu prüfen.</td></tr>
<tr><td>Der Anbieter muss die technischen und LZA-spezifischen Metadaten zum Datensatz hinzufügen.</td><td>Der Auftraggeber muss sich bei der Eingabe der Metadaten an die Vorgaben aus dem Handbuch halten. Sollten Fehler bei der Erstellung der Metadaten festgestellt werden, müssen diese innerhalb einer Woche an den Anbieter gemeldet werden.</td></tr>
<tr><td colspan="2">Die Metadaten müssen seitens des Auftraggebers zusammen mit dem digitalen Datensatz abgegeben werden.</td></tr>
<tr><td colspan="2">Sollten bei der Eingabe der Metadaten seitens des Anbieters Fehler unterlaufen sein, ist er dazu verpflichtet diese Fehler zu beheben. Sollte dies in einem Zeitraum von einer Woche nicht geschehen, ist der Anbieter dazu verpflichtet einen Strafe in Höhe von XX € zu zahlen.</td></tr>
<tr><td colspan="2">Dieses Service-Level ist teurer.</td></tr>
</table>

Tabelle 4: Service-Level: Metadaten II

<table>
<tr><th colspan="2">Metadaten III</th></tr>
<tr><td colspan="2">Der Produzent muss keine Metadaten zum Datensatz bereitstellen.</td></tr>
<tr><td colspan="2">Der Produzent der Daten gibt diese ohne jegliche Metadaten beim LZA-Anbieter ab. Der Anbieter ist im Folgenden für die Erstellung sämtlicher Metadaten zuständig.</td></tr>
<tr><td>-</td><td>Der Auftraggeber hat das Recht, die Metadaten auf Richtigkeit zu prüfen.</td></tr>
<tr><td>Der Anbieter muss alle notwendigen Metadaten zum Datensatz hinzufügen.</td><td>Sollten Fehler bei der Erstellung der Metadaten festgestellt werden, müssen diese innerhalb einer Woche an den Anbieter gemeldet werden.</td></tr>
<tr><td colspan="2">-</td></tr>
<tr><td colspan="2">Sollten bei der Eingabe der Metadaten seitens des Anbieters Fehler unterlaufen sein, ist er dazu verpflichtet diese Fehler zu beheben. Sollte dies in einem Zeitraum von einer Woche nicht geschehen, ist der Anbieter dazu verpflichtet einen Strafe in Höhe von XX € zu zahlen.</td></tr>
<tr><td colspan="2">Dieses Service-Level ist das teuerste.</td></tr>
</table>

Tabelle 5: Service-Level: Metadaten III

3.2.3 Finanzierung

„It is time for a financial review of the situation: we need to understand the full costs of curating data so that we can plan and budget for the required infrastructure more effectively. We need to construct new economic models to provide a robust foundation to funding plans, models which link research strategy and programme development, to operational support and infrastructure provision.“ (Lyon 2007, S. 53)

Die Finanzierung der Langzeitarchivierung ist seit langem eine ungeklärte Situation bei den Forschungsdaten. Bisher waren nach Beendigung eines Forschungsprojektes keine finanzielle Mittel seitens der Förderer zur Verfügung gestellt worden, um die Langzeitarchivierung der Daten zu finanzieren. Wie im Kapitel 2.3 aufgezeigt, tritt bei den Förderinstitutionen mittlerweile ein Umdenken in diesem Bereich ein, und es werden erste Empfehlungen formuliert, die auf die Nachhaltigkeit von Forschungsdaten abzielen. Als Folge der „OECD Principles and Guidelines for Access to Research Data from Public Funding“ (OECD 2007) bleibt zu hoffen, dass Grundsteine für den Aufbau von wissenschaftlichen Datenarchiven gelegt werden.

Die Finanzierung der Langzeitarchivierung kann in zwei verschiedene Abschnitte unterteilt werden. Dies sind zum einen die Anschubfinanzierung zum Aufbau eines Langzeitarchivs und zum anderen die Betriebsfinanzierung für den späteren dauerhaften Betrieb.

Die Anschubfinanzierung ist zu Beginn eines Vorhabens notwendig. Sie umfasst die Initial- und Beschaffungskosten, die als einmalige Kosten beim Start des Vorhabens auftreten. Dazu zählen beispielsweise die Informationsbeschaffung für ein passendes LZA-Systems und dessen Erwerbung im Anschluss sowie der Kauf von notwendigen Software-Lizenzen und die Anpassung von Software an die individuellen Vorgaben am Standort. Diese Finanzierung könnte möglicherweise im Rahmen von Forschungsprojekten erfolgen, die sich mit der Entwicklung eines technischen LZA-Systems beschäftigen (z.B. kopal). Auf dem zweiten KoLaWiss-Workshop sprach sich ein Vertreter der DFG dafür aus, mögliche Explorations-Projekte zu finanzieren, die Teilgebiete der Langzeitarchivierung von Forschungsdaten fachspezifisch untersuchen. Durch eine Vielzahl solcher Projekte und die damit verbundene Kooperation unter diesen Projekten kann ein Best-Practice-Szenario für Organisations- und Geschäftsmodelle für die Langzeitarchivierung von Forschungsdaten entstehen.

Darauf folgend ist der zweite Abschnitt der Finanzierung eine dauerhafte Betriebsfinanzierung der Langzeitarchivierung, die den ständigen Routinebetrieb des LZA-Dienstleisters gewährleistet. Betriebskosten fallen während des gesamten Betriebs der Langzeitarchivierung an in Form von Energie, Verwaltung und Infrastruktur. Hier sei noch einmal deutlich darauf hingewiesen, dass Kosten nicht nur für die technischen Systeme und Betrieb not-

wendig sind, sondern dass mindestens ebenso hohe Kosten bei organisatorischen Aufgaben der Langzeitarchivierung entstehen werden. Für diese Art der Finanzierung werden die Forschungsförderer nicht mehr zur Verfügung stehen, und es müssen neue Wege gefunden werden, die Finanzierung der anfallenden Kosten zu gewährleisten. Mögliche Ansätze werden im Folgenden untersucht.

Bei den beiden zuvor erwähnten KoLaWiss Workshops in Göttingen (vgl. Kapitel 2.2) wurden drei unterschiedliche Modelle für die Betriebsfinanzierung von Langzeitarchivierungslösungen von Forschungsdaten vorgeschlagen. Die ersten beiden Modelle gehen davon aus, dass bei der Beantragung eines Forschungsprojektes eine Pauschale für die Langzeitarchivierung der anfallenden Forschungsdaten, die während der Projektlaufzeit anfallen, zu Beginn mit beantragt werden. Diese „Datenpauschale“ wurde während der Diskussionen der Workshops mehrfach angesprochen und als sinnvoll erachtet. Auch ein anwesender Mitarbeiter einer Förderorganisation unterstützte die Idee der „Datenpauschale“, jedoch als eine Übergangssituation und nicht als eine dauerhafte Lösung des Problems, da er seine Institution nicht in der Rolle des Dauerfinanziers der Langzeitarchivierung von Forschungsdaten sieht. Das dritte Modell ist völlig unabhängig von dieser „Datenpauschale“.

Bei der „Datenpauschale“ wäre ein zweistufiges Verfahren eine mögliche Herangehensweise, bei dem bei der Beantragung des Projektes finanzielle Mittel für die Langzeitarchivierung der anfallenden Forschungsdaten im Rahmen der guten wissenschaftlichen Praxis mit beantragt werden. Damit können die Forschungsdaten für einen Zeitraum von zehn Jahren in ein Langzeitarchiv überführt werden. In einer zweiten Phase gegen Ende des Projektes wird evaluiert, inwiefern die bereits archivierten Daten noch für einen längeren Zeitraum vorgehalten werden sollen, und es werden weitere Forschungsdaten selektiert, die mit in das LZA-System überführt werden sollen.

Die bereitgestellte „Datenpauschale“ sollte in jedem Fall nur für ihren Bestimmungsgrund verwendet werden und darf im laufenden Projekt nicht auf andere Kostenstellen umgewidmet werden. Durch die Möglichkeit der Beantragung einer „Datenpauschale“ für die Langzeitarchivierung von Forschungsdaten zu Projektbeginn würden die Förderorganisationen ihren Empfehlungen nachgehen, sich mit der Langzeitarchivierung der anfallenden Forschungsdaten zu beschäftigen. Zudem wäre ein erster Schritt bei der Finanzierung der Langzeitarchivierung getan.

Bei der Finanzierung der Langzeitarchivierung von wissenschaftlichen digitalen Daten mit der „Datenpauschale“ wären die im Anschluss vorgestellten zwei Verfahren möglich.

Das erste mögliche Modell zur Finanzierung der Langzeitarchivierung von Forschungsdaten ist das „Kapitaldeckungsverfahren“. Dieses Verfahren wird bei der Finanzierung von Versicherungen (speziell der Altersvorsorge, aber

auch von Krankenversicherung, Pflegeversicherung und Arbeitslosenversicherung) verwendet.

Angewandt auf die Langzeitarchivierung würde dieses Verfahren darauf beruhen, dass Teile der Beiträge derjenigen, die Gelder für eine digitale Langzeitarchivierung von Forschungsdaten zahlen, auf dem Kapitalmarkt gewinnbringend angelegt werden. Die Idee hinter diesem Verfahren ist, dass das angelegte Kapital auf dem Finanzmarkt durch die Zinsen wächst, woraus im Anschluss die entstehenden Kosten für die Langzeitarchivierung der Forschungsdaten getragen werden können.

Dieses Verfahren setzt voraus, dass ein ausreichend hohes Kapital angelegt wird, damit die Langzeitarchivierungskosten von den Zinsen gedeckt werden können. Daher erscheint ein solches System eher für große Wissenschaftsstandorte in Frage zu kommen, die ausreichend Forschungsdaten in die Langzeitarchivierung überführen müssen. Möglich sind auch Kooperationen von Standorten, die sich für dieses Verfahren zusammenschließen, um am Markt gemeinsam aufzutreten.

Bei diesem Verfahren ist zu beachten, dass öffentliche Einrichtungen per Gesetz keinen Gewinn am Kapitalmarkt erwirtschaften dürfen. Dieser Punkt wurde auch bei den beiden KoLaWiss-Workshops als kritisch gesehen. Als Lösung wäre eine Gründung einer privatwirtschaftlich ausgerichteten Institution denkbar, die beispielsweise als eine GmbH oder eine GbR als Mittler auftreten würde. Sie würde lediglich die Rolle der Finanzierung für die Langzeitarchivierung übernehmen.

Vor dem Hintergrund der aktuellen Weltwirtschaftskrise erscheint es jedoch fraglich, ob es sinnvoll wäre, die Finanzierung der Langzeitarchivierung von wichtigen Forschungsdaten vom derzeitigen Kapitalmarkt abhängig zu machen.

Ein zweites Verfahren ist das „Umlageverfahren", welches ursprünglich ein Verfahren zur Finanzierung von Sozialversicherungen (z.B. Altersvorsorge, Krankenversicherung und Arbeitslosenversicherung) darstellt.

Innerhalb der Finanzierung der Langzeitarchivierung von Forschungsdaten kann das „Umlageverfahren" so angewandt werden, dass eingezahlte Beiträge unmittelbar für die Finanzierung der digitalen Langzeitarchivierung verwendet werden. Diese Beträge werden vermutlich ausschließlich für die Betriebskosten der anfallenden Kosten ausreichen und können keine großen Beschaffungskosten decken. Durch dieses Verfahren decken die jeweils aktuellen Projekte die Kosten der gesamten im Langzeitarchiv befindlichen Forschungsdaten.

Diese Methode kann auch von den öffentlichen Einrichtungen ohne rechtliche Bedenken durchgeführt werden. Kritisch bei diesem Vorgehen ist, dass immer ausreichend neue Forschungsprojekte ihre Daten in das Archiv überfüh-

ren, damit ständig neue Einnahmen für den LZA-Dienstleister vorhanden sind. Weiterhin ist problematisch, dass keine finanziellen Rücklagen zur Verfügung stehen, falls kostenintensive technische Anschaffungen getätigt werden müssen. Finanzielle Engpässe können entstehen, falls in einem Zeitraum kaum Forschungsdaten zur Langzeitarchivierung in das Archiv abgegeben werden, da dadurch auch keine Beiträge erwirtschaftet werden.

Als letzte mögliche Methode für die Finanzierung wird das „Steuermodell" gesehen, welches in vielen öffentlichen Einrichtungen Anwendung findet, indem Steuern für Dienstleistungen erhoben werden, um diese der Allgemeinheit zur Verfügung zu stellen. Dieses Modell der Finanzierung würde auch ohne die zuvor erwähnte „Datenpauschale" auskommen.

Bei der Finanzierung der Langzeitarchivierung von Forschungsdaten würde die Anwendung so aussehen, dass die notwendigen finanziellen Mittel durch Steuern erhoben werden. Dadurch würde jeder Bürger seinen Beitrag an der Finanzierung der dauerhaften Nutzbarkeit von Forschungsdaten beitragen.

Von Vorteil bei diesem Modell ist, dass die Einnahmen durch die Steuern relativ gut zu kalkulieren sind, da sie nicht unmittelbar von der Produktion der Forschungsdaten abhängig sind. Als problematisch kann die derzeitige Diskussion um Steuersenkungen und der hoch verschuldete Staatshaushalt angesehen werden. Vermutlich wird sich kein Politiker und auch nur wenige Steuerzahler für die Finanzierung der Langzeitarchivierung von Forschungsdaten interessieren.

Zusammenfassung

Vermutlich wird sich keins der zuvor genannten Modelle alleinig durchsetzen können, da sie alle noch nicht an die notwendigen Anforderungen angepasst sind. Voraussichtlich wird es ein Mischmodell aus den zuvor genannten Modellen geben, das den Anforderungen entsprechen kann. So ist beispielsweise denkbar, dass eine Grundfinanzierung im Rahmen des Steuermodells erfolgt, damit eine dauerhafte Absicherung der Finanzierung gegeben ist. Zusätzlich wird eine Mischung aus dem Umlage- und dem Kapitaldeckungsverfahren durchgeführt, damit nicht die Risiken des Kapitalmarktes in vollem Umfang zur Geltung kommen können.

3.3 Vorhandene Geschäftsmodelle für die Langzeitarchivierung

Bis heute liegen keine umfangreichen Geschäftsmodelle für die Langzeitarchivierung von Forschungsdaten vor. Erste brauchbare Lösungsansätze sind innerhalb von Projekten erfolgt und sollen im Folgenden näher betrachtet werden. Beide der folgenden Ansätze beziehen sich nicht auf die Langzeitarchivierung von Forschungsdaten, sondern zeigen Beispiele, wie ein Geschäftsmodell umgesetzt werden kann. Zum einem wird kopal betrachtet, das eher eine technische Lösung der Langzeitarchivierung darstellt und zum an-

deren wird Portico betrachtet, das zwar zurzeit keine Forschungsdatenarchivierung anbietet, aber vom Konzept als interessant erscheint.

3.3.1 kopal

Das Projekt kopal[28] ist ein dreijähriges BMBF-gefördertes Forschungsprojekt (2004-2007). Ziel des Projektes war es, ein kooperatives Langzeitarchiv für digitale Objekte praktisch zu erproben und im Anschluss technisch umzusetzen. Die beteiligten Projektpartner sind unter der Leitung der Deutschen Nationalbibliothek (DNB) die SUB Göttingen, die IBM Deutschland GmbH und die GWDG.

Zum Ende der Projektlaufzeit konnte das Projekt auf Grund von mehrfachen Nachfragen von interessierten Einrichtungen ein Nutzungskonzept für das System vorlegen und damit auch konkrete Zahlen nennen, die bei einer Nutzung anfallen.

kopal ist ein frei konfigurierbare s und auf einzelnen Anforderungen anpassbares System. Die Modularität ermöglicht es den Nutzern des Systems, je nach ihren Bedürfnissen Erweiterungen vorzunehmen, um sie dem speziellen Workflow anzupassen. Dabei übernehmen die Projektpartner verschiedene Rollen beim Anbieten von Dienstleistungen, Systemen und Lizenzen (siehe Abbildung 11).

Abbildung 11: Ansprechpartner im kopal-Workflow (Quelle: kopal 2007).

Das kopal-System besteht aus zwei Hauptkomponenten, die unterschiedliche Vermarktungswege gehen. Das *Digital Information Archiving System* (DIAS) ist eine Entwicklung des Projektpartners IBM, der dieses System an die Anforderungen von kopal angepasst hat und es so unter anderem mandantenfähig gemacht hat. Für dieses System fallen Lizenzkosten bei IBM an. Die zweite Hauptkomponente ist die Open-Source-Softwarebibliothek *kopal Library for Retrieval and Ingest* (koLibRI), die das Erstellen, Einspielen

[28] http://kopal.langzeitarchivierung.de/

und Abfragen von Archivpaketen[29] sowie künftig Maßnahmen der Langzeiterhaltung wie Datenformatmigrationen unterstützt. Für dieses System fallen im Grunde keine Kosten an, da die Software Open-Source ist. Lediglich mögliche Anpassungen des Moduls an die spezifischen Anforderungen können finanzielle Aufwendungen notwendig werden lassen.

Damit konkrete Zahlen für das kopal-System genannt werden können, wird die mögliche Inanspruchnahme des Systems in drei Nutzungsmodelle unterteilt (vgl. im Folgenden kopal 2007):

- „kopal-Teilnehmer"
  Bei diesem Nutzungsmodell steht dem Auftraggeber selbst kein eigenes Archiv vor Ort zur Verfügung. Die Daten werden kommissarisch durch einen anderen kopal-Mandanten archiviert. Die Institution betreibt lediglich das koLibRI-Tool selbst und kann somit die Daten einspeisen, verwalten und darauf zugreifen.
- „kopal-Mandant"
  Die Institution, die diese Nutzungsart wählt, hat bereits mehr Einfluss auf das eigentliche System, das jedoch nicht bei der Institution selbst betrieben wird. Die Institution kann auf eine kopal-Instanz bei einem anderen Service-Anbieter zugreifen.
- „kopal-Eigenbetrieb"
  Bei dieser Nutzungsart läuft eine komplett eigenständige kopal-Installation bei der Institution. Bei der Einrichtung und dem Betrieb des Systems erfolgt eine unterstützende Tätigkeit durch das kopal-Projektteam.

Die in kopal anfallenden Kosten teilen sich in verschiedene Kostenfaktoren auf, die wiederum von verschiedenen Merkmalen beeinflusst werden. Kostenfaktoren sind dabei die Erstberatung durch Projektmitarbeiter, die für die interessierten Institutionen kostenlos ist, sowie Projektplanung und Beratung. Weiterhin fallen Kosten für die Hardware und für Lizenzen für das DIAS-System an. Zudem müssen institutions-spezifische Anpassungen der Open-Source-Software koLibRI vorgenommen werden. Einer der vermutlich größten Kostenfaktoren (zumindest über die Zeit gesehen) ist der folgende Dauerbetrieb des Systems.

Diese Kosten sind von verschiedenen Merkmalen abhängig, wodurch je nach Institution unterschiedliche Kalkulationen entstehen. Dies sind auf der einen Seite die Menge, Heterogenität und Komplexität der zu archivierenden digitalen Objekte und ihrer Metadaten. Auf der anderen Seite sind auch die gewünschten Zugriffsmöglichkeiten und Schnittstellen seitens der Institution ein Merkmal bei der Berechnung der Kosten.

---

[29] Gespeicherte Objekte mit ihren Metadaten.

Das Projekt stellt zum Ende seiner ersten Projektphase eine exemplarische Kostenkalkulation vor, bei der es Kosten für die drei möglichen Nutzungsmodelle vorstellt.

- „kopal-Teilnehmer“
  Die Institution betreibt die kostenfreie koLibRI-Software selbst und kauft notwendige Servicekomponenten, wie die Installation der Software und notwendige Beratung, bei den kopal-Partnern ein. Bei einer einfachen Installation ist mit einem Aufwand von 10 Personentagen zu kalkulieren. Die Kosten der kommissarischen Archivierung hängen ferner von der Menge der zu archivierenden Daten ab.
- „kopal-Mandant“
  Die notwendigen DIAS-Servicekomponenten werden vereinbart, dabei fallen für Lizenzkosten je nach Größe des Systems Kosten von ca. 96.000 bis 385.000 € an. Das koLibRI-System ist wiederum frei zu verwenden, es ist lediglich eine Auswahl zu treffen, welche zusätzlichen Komponenten selbst oder durch Zukauf entwickelt werden. Weiterhin fallen Kosten bei der Inanspruchnahme von Beratungen an. Die Beratungen im Rahmen des DIAS-Systems können dabei sehr unterschiedlich ausfallen. Hier sind zweistündige Kurzeinführungen bis hin zu mehrtägigen Workshops denkbar und je nach Anforderungen des Auftraggebers buchbar. Für einen zweitägigen Workshop beispielsweise, der die Zielsetzung „Vorstellung der DIAS-Software und Diskussion technischer Fragen“ verfolgt, werden Kosten von ca. 5.000 € anfallen. Abschließend entstehen noch Kosten für die Mitnutzung eines vorhandenen Systems als kopal-Mandant, die je nach Beratungs- und Nutzungsumfang jährliche Kosten von 50.000 bis 200.000 € erzeugen. Diese Berechnungen sind von mehreren Merkmalen wie Datenmenge, Zugriffsanforderungen, usw. abhängig.
- „kopal-Eigenbetrieb“
  Das gesamte kopal-Archivsystem läuft in der Institution und wird durch Zukauf von Beratung und ggf. notwendiger Entwicklung am System eigenständig betrieben. Die Investition für Hard- und Software beträgt für ein mittleres System etwa 750.000 €. Diese Kosten teilen sich in 40% Softwarelizenzen und 60% Systembereitstellung und -betrieb auf. Zusätzlich fallen wie zuvor genannt Kosten für Beratungen und Schulungen an.

3.3.2 Portico

Im Rahmen von Portico wird das Ziel verfolgt, digitale wissenschaftliche Literatur auf Dauer zu archivieren. Dabei beschränkt sich der Dienst auf wissenschaftliche Zeitschriften (Electronic Scholarly Journals), die ein Peer-Review Verfahren durchlaufen haben. Portico startete 2002 als ein Projekt gefördert durch JSTOR und mit Unterstützung der Andrew W. Mellon Foundation. Seit

2005 ist Portico in eine gemeinnützige Organisation überführt wurden[30]. Portico bietet seine Dienstleistungen Bibliotheken und Verlagen an.

Die Bibliotheken verfolgen dabei das Ziel, den dauerhaften Zugriff auf die digitalen Ressourcen sicherzustellen, die sie erworben bzw. lizenziert haben. Häufig sind die Bibliotheken jedoch nicht im Besitz der Ressourcen, da der Verlag die Archivierung bzw. das Hosten seiner Zeitschriften übernimmt. Ein Problem kann beispielsweise auftreten, wenn die Bibliotheken das Zeitschriftenabo kündigen und der Verlag den Zugriff auf die alten Ausgaben nicht mehr gewährt. Schwierigkeiten können sich zudem ergeben, wenn beim Verleger technische Probleme auftreten und die Bibliothek dadurch nicht mehr auf die Zeitschriften zugreifen kann.

Portico garantiert allen Bibliotheken, die bei ihm Mitglied sind, den campusweiten Zugriff auf die archivierten Fachzeitschriften bei Eintritt eines sog. „trigger events“. Unter diesen Events versteht Portico die folgenden Fälle:

- Verlag stellt den Betrieb ein
- Verlag stellt die Veröffentlichung einer Zeitschrift ein
- Verlag erlaubt keinen Zugriff mehr auf die früheren Ausgaben einer Zeitschrift
- Katastrophenfall sowie ein dauerhafter Ausfall der technischen Bereitstellung durch den Verlag

Dadurch müssen Bibliotheken zumindest für den von Portico archivierten Teil ihres Bestandes keine eigene Langzeitarchivierungsstrategien entwickeln und umsetzen. Zudem könnten sie auf die Print-Bestände im eigenen Archiv verzichten, die bei Portico integriert sind.

Als Nachteil kann gesehen werden, dass Portico ausschließlich den intellektuellen Content des digitalen Objektes sichert, nicht jedoch die publikationsspezifische Erscheinungsform (Look & Feel). Auch die Funktionalität der Publikation in ihrer Anwendungsumgebung wird nicht archiviert. Zudem ist Portico zurzeit auf noch sehr spezielle Materialarten festgelegt: So ist die Langzeitarchivierung von Forschungsdaten nicht möglich.

Die Kosten (Annual Archive Support) für Bibliotheken hängen dabei vom jährlichen Beschaffungsetat der Institution ab. Sie starten bei einem Bewerbungsetat der Bibliothek von weniger als 150.000 US $ mit 1% des jeweiligen Etats und steigen an bis zu einem Wert von 24.000 US $ bei einem Bewerbungsetat von 25-30 Millionen US $.[31]

---

[30] Mit Unterstützung durch JSTOR, Ithaka, The Library of Congress, and The Andrew W. Mellon Foundation.

[31] http://www.portico.org/publishers/pub_contribution.html

Die Verleger haben durch Portico die Vorteile, dass sie nicht selbst für die Langzeitarchivierung ihrer Titel sorgen müssen und damit den Bibliotheken einen dauerhaften Zugriff auf ihre Publikationen anbieten können. Die Verlage gehen dazu mit Portico einen Lizenzvertrag ein und zahlen einen jährlichen Beitrag, der jedoch in seiner Höhe nicht weiter definiert wird.

Zusammenfassung

Bei kopal können die drei angegebenen Nutzungsmodelle im weitesten Sinn als Service-Level verstanden werden. Ansonsten bietet kopal keine detaillierte Kostenaufteilung an, zumindest nicht auf den ersten Blick. Vermutlich muss mit den Projektpartnern Kontakt aufgenommen werden und ein genaues Szenario beschrieben werden, um eine genauere Kostenberechnung zu erhalten.

Portico ist mittlerweile ein System, das von vielen Institutionen genutzt wird und in dem eine Vielzahl von e-Journals integriert sind. Derzeit nehmen 490 Bibliotheken und 71 Verleger an Portico teil. Das Angebot umfasst mehr als 12 Mio. Artikel und mehr als 7.000 Titel.

	kopal	Portico
Vorteile	• Modulares Konzept, das einfach erweiterbar ist • Zum Teil Open-Source Software • Konzept zum Preservation Planning • Unterstützt die Formatmigration • Keine Formatbeschränkung • Keine Mengenbeschränkung	• Unterstützt durch die Library of Concres und die Mellon W. Foundation • Backupsystem für Bibliotheken, im Falle von „trigger events“
Nachteile	• Hohe Anfangsinvestition notwendig • Lizenzkosten	• Beschränkung auf einzelne Materialarten (e-Journals, e-Books, e-newspapers) • Es entstehen in jedem Fall Kosten • Die Artikel werden nicht in ihrem Original archiviert, daher kann sich das Look & Feel ändern

Tabelle 6: Vor- und Nachteile von kopal und Portico.

Als ein mögliches Geschäftsmodell sollte auch die Auslagerung an einen externen Anbieter berücksichtigt werden. Dies könnte der zuvor genannte LZA-

Knoten sein, bei dem die Dienstleistung „Langzeitarchivierung von Forschungsdaten“ eingekauft werden, oder aber auch ein kommerzielle Anbieter, der die Langzeitarchivierung der wissenschaftlichen digitalen Daten übernimmt. Noch sind weder die öffentlichen Einrichtungen noch die kommerziellen Anbieter in der Lage, ein Angebot für die sichere Langzeitarchivierung für Forschungsdaten anbieten zu können. Eine Interessante Aussage eines kommerziellen Anbieters von LZA-Systemen in einer Diskussion mit einem CEO eines großen Industrieunternehmers war: „Langfristige Ziele entsprechen nicht dem Aktionärs- oder Managerinteressen. Darauf bin ich nicht gegoalt!“[32]. Dies zeigt auch, dass zurzeit in der Industrie noch keine große Nachfrage nach Langzeitarchivierungslösungen vorhanden ist, und dass auch in diesen Bereichen ein Umdenken stattfinden muss.

3.4 Zusammenfassung: Notwendige Aspekte für die Entwicklung eines Geschäftsmodells für die Langzeitarchivierung

Es wird vermutlich noch ein weiter Schritt hin zu einem ausgereiften Geschäftsmodells für die Langzeitarchivierung von Forschungsdaten sein, aber die zuvor gezeigten Ansätze sind in vielen Punkten sehr vielversprechend. Es bleibt abzuwarten inwiefern die Projekte durch ihre Forschungstätigkeiten dem Ziel ein wenig näher kommen und inwiefern die Förderorganisationen mit ihren Empfehlungen dem Ganzen dabei unterstützend tätig werden können. Letzten Endes ist auch ein Umdenken in den Köpfen der Wissenschaftler notwendig, die ihre Forschungsergebnisse auch durch eigenes Interesse in ein Langzeitarchiv überführen möchten.

Damit eine Langzeitarchivierung von Forschungsdaten organisatorisch möglich ist, müssen die in diesem Kapitel aufgezeigten Aspekte erfüllt werden. Dazu zählt die Bestimmung der beteiligten Akteure an der Langzeitarchivierung von Forschungsdaten, eine Kostenberechnung für die anfallenden Daten, das Aufstellen von Service-Level-Agreements der angebotenen Dienste sowie die Klärung der notwendigen Finanzierung.

---

[32] Berichtet auf dem 2. KoLaWiss-Workshop.

4 Schluss

4.1 Fazit

Die vorliegende Arbeit zeigt, dass zur Entwicklung eines Geschäftsmodells für die Langzeitarchivierung von Forschungsdaten viele Voraussetzungen zu erfüllen sind. Die Umfrage an den Standorten Göttingen und Zürich hat gezeigt (vgl. Kapitel 2.1), dass großes Interesse an Langzeitarchivierung besteht, aber kaum Planungen für eine konkrete Umsetzung vorliegen. Dieses Ergebnis verdeutlicht, dass Handlungsbedarf zu einer konkreten Herangehensweise fehlt und dass in diesem Bereich Unterstützung stattfinden muss.

Zu Beginn der Entwicklung eines Geschäftsmodells für die Langzeitarchivierung von Forschungsdaten bedarf es einer umfangreichen Evaluation der vorliegenden Daten am Standort, um sich ein Bild von Datenmenge, Heterogenität der Datenformate und den bereits vorhandenen Strategien zu machen. In einem nächsten Schritt müssen nationale und internationale Empfehlungen sowie Policies berücksichtigt werden. Die zuvor genannte Umfrage hat auch gezeigt, dass viele lokale Wissenschaftler kaum Informationen über das Vorhandensein solcher Policies haben. Hier besteht Handlungsbedarf, diese Informationen am Wissenschaftsstandort unter den relevanten Forschern bekannt zu machen.

Für die Umsetzung eines Geschäftsmodells am Wissenschaftsstandort müssen die Akteure und deren Rolle im Prozess der Langzeitarchivierung identifiziert und deren vorhandene Kompetenzen berücksichtigt werden. Dieses Vorgehen ist eine Grundvoraussetzung bei der Umsetzung eines Geschäftsmodells (vgl. Kapitel 3.1). Nachdem die zugrundeliegenden Daten in einer Evaluation am Standort erhoben wurden und die beteiligten Akteure identifiziert worden sind, können die Kosten anhand von Modellen (vgl. Kapitel 3.2.1) berechnet werden. Diese Kosten können von verschiedenen Merkmalen wie Menge der anfallenden Daten oder auch deren Datei-Formaten abhängen. Im Anschluss müssen Sevice-Level-Agreements erarbeitet werden, die eine vertragliche Regelung zwischen einem LZA-Dienstleister und dem Auftraggeber in ihren Leistungen näher beschreibt. Diese SLA helfen bei der transparenten Darstellung dieser Leistungen des Anbieters, und durch die Auswahl verschiedener SL kann dem Auftraggeber ein breites Angebot an unterschiedlichen Leistungen angeboten werden. Erste Geschäftsmodelle für die Langzeitarchivierung liegen bereits vor (vgl. Kapitel 3.3). Bei diesen handelt es sich jedoch nicht um Stragtegien für die Langzeitarchivierung von Forschungsdaten. Das vorgestellte kopal-Projekt eignet sich noch am ehesten für die Materialart *Forschungsdaten*, da es sich nicht auf Datei-Formate beschränkt, ist aber größtenteils eine technische Umsetzung und noch nicht ein komplett selbstständiges System für die Langzeitarchivierung von Forschungsdaten. Portico bietet seinen Service zurzeit Bibliotheken für die Sicherung von e-Journals und anderen digitalen Publikationen an und ist zur-

zeit noch nicht für Forschungsdaten geeignet, jedoch ist das Geschäftsmodell bereits etabliert, was die Beteiligung der zahlreichen Bibliotheken und Verlage zeigt, und könnte möglicherweise für die Langzeitarchivierung von Forschungsdaten adaptiert werden.

4.2 Empfehlungen für die Entwicklung von Geschäftsmodellen für die Langzeitarchivierung von Forschungsdaten

Im Folgenden werden als Ergebnis der vorliegenden Arbeit Empfehlungen gegeben, die bei der Konzeption eines Geschäftsmodells für die Langzeitarchivierung von Forschungsdaten bedacht werden müssen:

1. Forschung und Entwicklung im Bereich der Langzeitarchivierung von Forschungsdaten muss weiter betrieben werden: Dies wird vermutlich noch einige Zeit dauern, und es werden auch nicht gleich alle Ansätze zum richtigen Ziel führen. Besonders bei einem Thema wie der Langzeitarchivierung, bei dem erst nach langer Zeit feststellt werden kann, ob der gewählte Weg erfolgreich war, ist die Suche nach den richtigen Ansätzen bisweilen sehr schwer. Vorhandene Strategien müssen somit stetig überprüft und weiterentwickelt werden.
2. Pilotprojekte mit fachspezifischen Anforderungen an die Langzeitarchivierung von Forschungsdaten müssen initiiert werden: Anhand von Pilotprojekten sollen an konkreten Beispielen an Standorten Geschäftsmodelle und organisatorische Abläufe erarbeitet werden, die im Folgenden auf andere Standorte Anwendung finden können. Diese Anforderung ist seitens der Forschungsförderer auch in dem Papier der DFG (DFG 2009, S. 2) wiederzufinden und wurde von dem DFG-Vertreter auf dem 2. KoLaWiss Workshop als gangbarer Weg beschrieben. In dem konkreten Beispiel müssen Finanzierungsmöglichkeiten für die Archivierung der Daten gefunden werden und detaillierte Service-Level für die vor Ort anfallenden Daten entwickelt werden. Die notwendigen Akteure (vgl. Kapitel 3.1) vor Ort müssen identifiziert und kontaktiert werden, um mit ihnen den weiteren Ablauf zu besprechen.
3. Langzeitarchivierung von Forschungsdaten benötigt ausgebildetes Personal: Das Personal, das heute mit der Aufgabe der Langzeitarchivierung betreut ist, ist größtenteils nicht für diese speziellen Tätigkeiten ausgebildet. Darüber hinaus gibt es kaum Lösungen, wie eine Weitergabe der Kompetenzen unter den Kollegen erfolgen soll. Für eine effiziente Langzeitarchivierung der wissenschaftlichen Daten ist hingegen sowohl fachlich als auch technisch ausgebildetes Personal notwendig. Es wird daher die Einrichtung eines *Preservation Officer* empfohlen, der ausgewiesene Kompetenzen auf den Gebieten der Langzeitarchivierung und dem Umgang mit Forschungsdaten hat. Diese Person ist nicht nur an Standorten vertreten, an denen sich ein LZA-System selbst befindet, sondern muss an allen Standorten zur festen Belegschaft gehören. Der *Preservation Officer*

ist dort die erste Anlaufstelle, wenn es um Fragen der Langzeitarchivierung von Forschungsdaten geht und stellt ein Bindeglied zwischen den an der Langzeitarchivierung beteiligten Dienstleistern dar. Im Vergleich zu anderen Ländern existiert in der Bundesrepublik an wissenschaftlichen Bibliotheken das Konzept der Fachreferenten, das bereits etabliert ist. Diesen Vorteil gilt es in Zukunft zu nutzen, die Fachreferenten für das Thema Langzeitarchivierung zu interessieren und im Anschluss dazu auszubilden. Dadurch kann auf eine funktionierende organisatorische Struktur mit existierenden Verantwortlichkeiten zurückgegriffen werden.

4. Weiter-, Fort- und Ausbildung auf dem Bereich der Langzeitarchivierung von Forschungsdaten ist dringend notwendig: Pfeiffenberger stellt sehr treffend fest, dass „der Flaschenhals [...] das disziplinspezifische kompetente Personal sein [wird].“ (Pfeiffenberger 2007, S. 209) Daher sind Weiter-, Fort- und Ausbildung im Bereich der Langzeitarchivierung – und dort auch im Speziellen für den Bereich der Forschungsdaten – dringend notwendig. Hier sollte von den Ergebnissen profitiert werden, die Projekte wie nestor – das Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen in Deutschland[33] und das europäische Projekt DigitalPreservationEurope(DPE)[34] im Bereich Training erzielen.
5. Geschäftsmodelle für Langzeitarchivierung von Forschungsdaten müssen entwickelt werden: An den Standorten, an denen wissenschaftliche Daten bei der Forschung anfallen, müssen sich Einrichtungen mit der Entwicklung eines Geschäftsmodells zu deren Langzeitarchivierung beschäftigen. Dazu sind Service-Level-Agreements zu beschreiben, damit eine transparente Vertragsituation geschaffen werden kann. Zudem muss im Vorfeld eine Berechnung der anfallenden Kosten für die zu archivierenden Daten erstellt werden. Dazu sind nach LIFE (vgl. Kapitel 3.2.1) die Kosten für die Lebenszyklus-Stufen des digitalen Dokumentes und die organisatorischen Kosten (non Lifecycle-Cost) zu berechnen. Auf Grundlage dieser Vorgaben können dann Möglicheiten der Finanzierung der Langzeitarchivierung geprüft werden.
6. Langzeitarchivierung von Forschungsdaten wird nur als Kooperation der in Kapitel 3.1 genannten Akteure möglich sein: In Zukunft wird die Langzeitarchivierung von Forschungsdaten sowie auch die anderer Datenarten in Form von Kooperation erfolgen müssen. Es wird kaum Institutionen geben, die diese umfangreiche und kostenintensive Aufgabe allein erfüllen können wird. Dies zeigt beispielsweise die Kostenaufstellung vom kopal-Projekt, die verdeutlicht, dass es wesentlich wirtschaftlicher

[33] http://www.langzeitarchivierung.de
[34] http://www.digitalpreservationeurope.eu/

ist, wenn Wissenschaftsstandorte ein gemeinsames kopal-System betreiben. Dies erscheint umso ratsamer, da die Kooperationen bereits bei den Nutzungsmodellen in Form von kopal-Teilnehmern und kopal-Mandanten bedacht wurde. Die Standorte, die die LZA-Systeme betreiben werden, werden vermutlich große Wissenschaftsstandorte sein, an denen zurzeit bereits Kompetenzen im Bereich der Langzeitarchivierung vorhanden sind, sei es durch gesetzliche Sammelaufträge oder durch umfangreiche Beteiligungen an Forschungsprojekten in diesem Bereich, wie das Beispiel Göttingen zeigt. Dadurch könnten in Zukunft sog. „LZA-Knoten" im Sinne von Kompetenzzentren entstehen, die zum Einem die Langzeitarchivierung der vor Ort anfallenden Daten übernehmen werden und zum Anderen als Dienstleister für weitere Standorte auftreten können.

4.3 Ausblick

In Zukunft werden sich die Akteure der Langzeitarchive mit der Frage beschäftigen müssen: Wie kommen die Langzeitarchive an die Daten, die archiviert werden sollen? Dieses Problem, nämlich die Wissenschaftler davon zu überzeugen und dazu zu motivieren, ihre Forschungsdaten in ein digitales Langzeitarchiv zu überführen und dabei auch die notwendigen Vorarbeiten, wie beispielsweise die Erstellung von deskriptiven Metadaten zu erledigen, muss bereits jetzt angegangen werden.

Es scheint zudem paradox, dass die Forschung auf dem Gebiet der Langzeitarchivierung größtenteils im Rahmen von zeitlich begrenzten Drittmittel-Projekten geleistet wird. Ein Dauerbetrieb – wie er sinnvoller Weise für ein Langzeitarchiv umgesetzt werden muss – sollte jedoch nicht von befristetem Personal aufrecht erhalten werden, sondern muss in die Infrastruktur einer oder auch mehrerer Institution überführt werden. Dies gilt auch für alle weiteren Ressourcen, die für einen reibungslosen Ablauf notwendig sind. Hier wird in nächster Zukunft ein Umdenken bei den Förderern, aber auch in der Politik, sei es national oder auch international, stattfinden müssen.

Die Wissenschaftler am Forschungsstandort benötigen für die Überführung ihrer Forschungsdaten lokale Unterstützung. Dem Endnutzer dieser Daten sollte ein globaler und einheitlicher Zugriff möglich sein, da ein Nutzer von Forschungsdaten nicht erst in jedem Langzeitarchiv suchen möchte, ob die Daten dort vorhanden sind. Bei diesem Zugriff muss beachtet werden, dass die Nutzer gewisse Erwartungen haben: So werden einfache Suchanfragen (wie bei Google) auch auf wissenschaftlicher Ebene oft umfangreicheren Zugriffsmöglichkeiten vorgezogen. Diesen Forderungen muss nachgekommen werden. Die Schwierigkeit dabei liegt darin, dem zukünftigen Nutzer einen adäquaten Zugriff zu gewährleisten, obwohl seine Anforderungen heute noch nicht klar definiert sind. In diesem Zwischenspiel von lokaler Erstellung der Daten und dem globalen Zugriff müssen die Rollen beleuchtet werden, damit Standards festgelegt werden, die für einen solchen Austausch unabdingbar sind.

In der Literatur, Mailinglisten und auf verschiedenen Blogs wird das Thema der Langzeitarchivierung von Forschungsdaten engagiert diskutiert. Zudem hat die interessierte sowie zahlreiche Beteiligung und die anschließende rege Diskussion bei den beiden KoLaWiss-Workshops gezeigt, dass die Entwicklung von Geschäftsmodellen und deren konkrete Umsetzung ein Problem ist, das zwar weitgehend identifiziert, jedoch bei weitem noch nicht als gelöst betrachtet werden kann.

Es bleibt abzuwarten, wie sich die Förderinstitutionen und die Politik mit diesem Thema in Zukunft befassen und ob Pilotprojekte Ergebnisse für eine Umsetzung aufzeigen können.

Literaturverzeichnis

Allianz-Initiative Digitale Information (2008): *Schwerpunktinitiative »Digitale Information« der Allianz-Partnerorganisationen*. Berlin, 2008. http://www.dfg.de/aktuelles_presse/das_neueste/download/pm_allianz_digitale_information_details_080612.pdf [20.05.09]

Ashley, Kevin (1999): *Digital Archive Costs: Facts and Fallacies*. DLM-Forum on Electronic Records (DLM '99). Brüssel, DLM-Forum, 1999. http://ec.europa.eu/archives/ISPO/dlm/fulltext/full_ashl_en.htm [20.05.09]

Ayris, Paul et al. (2008): *The LIFE2 Final Projekt Report*. 2008 http://eprints.ucl.ac.uk/11758/1/11758.pdf [20.05.09]

Beargrie, Neil / Chruszcz, Julia /Lavoie, Brian (2008): *Keeping Research Data Safe. A cost model and guidance for UK universities*. 2008. http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf [20.05.09]

Berliner Erklärung 2003: *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*. Berlin, 2003. http://oa.mpg.de/openaccess-berlin/berlindeclaration.html [20.05.09]

Björk, Bo-Christer (2007): *Evaluation of the Costing Activities and Economic Models for digital Curation using the LIFE Methodology*. LIFE Project, 2007. http://eprints.ucl.ac.uk/7684/1/7684.pdf [20.05.09]

Borghoff, Uwe M. (2005): *Vergleich bestehender Archivierungssysteme*. nestor Materialien 3, 2005. http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-20050117016 [20.05.09]

Boutsiouci, Pascalia / Metje, Daniel (2008): *Digitale Langzeitarchivierung - Erhebung an der ETH Zürich und der Georg-August-Universität Göttingen.* Vortrag, gehalten auf dem 97. Deutschen Bibliothekartag in Mannheim, am 05. Juni 2008. http://www.opus-bayern.de/bib-info/volltexte/2008/556/ [20.05.09]

Brase, Jan (2004): Using Digital Library Techniques - Registration of Scientific Primary Data. In: Jones M. / Fox, E.A. / Shen R. (Hrsg.): *Research and Advanced Technology for Digital Libraries. Lecture Notes in Computer Science.* Springer-Verlag, Heidelberg, 2004. S. 488-494. http://www.kbs.uni-hannover.de/Arbeiten/Publikationen/2004/brase_TIB_hannover.pdf [20.05.09]

CCSDS Secretariat (2002): *Reference Model for an Open Archival Information System (OAIS)*. CCSDS 650.0-B-1, Blue Book, 2002. http://public.ccsds.org/publications/archive/650x0b1.pdf [20.05.09]

Deutsche Forschungsgemeinschaft (DFG) (1998): *Vorschläge zur Sicherung guter wissenschaftlicher Praxis*. Empfehlungen der Kommission "Selbstkontrolle in der Wissenschaft". Denkschrift. Wiley-VCH Verlag, Weinheim, 1998. http://www.dfg.de/aktuelles_presse/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf [20.05.09]

Deutsche Forschungsgemeinschaft (DFG) (2007): *Elektronische Publikationen. Aktionslinie „Entwicklung von Organisations- und Geschäftsmodellen zur Langzeitarchivierung"*. Merkblatt. Bonn, 2007. http://www.dfg.de/forschungsfoerderung/formulare/download/12_113.pdf [20.05.09]

Deutsche Forschungsgemeinschaft (DFG) (2009): *Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten*. Bonn, 2009. http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/veroeffentlichungen/dokumentationen/download/ua_inf_empfehlungen_200901.pdf [20.05.09]

Digital Preservation Coalition (DPC) (2008): *The Preservation Management of Digital Material Handbook.* www.dpconline.org/graphics/handbook/ [20.05.09]

Funk, Stefan (2008): Digitale Objekte. In: Neuroth, Heike et al. (Hrsg.): *nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Version 1.5. Göttingen, 2008, S. 9-3 bis 9-6. http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_236.pdf [20.05.09]

Klump, Jens (2008a): *Anforderungen von e-Science und Grid-Technologie an die Archivierung wissenschaftlicher Daten*. nestor Materialien 9, 2008. http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-2008040103 [20.05.09]

Klump, Jens (2008b): Wissenschaftliche Primärdaten. In: Neuroth, Heike et al. (Hrsg.): *nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Version 1.5. Göttingen, 2008, S.15-54 bis 15-63. http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_275.pdf [20.05.09]

Klump, Jens et al. (2006): Data publication in the Open Access Initiative. In: *Data Science Journal*. Volume 5, 2006, S.79-83. http://www.jstage.jst.go.jp/article/dsj/5/0/79/_pdf [20.05.09]

Lang, Susanne (2008): *Ist-Zustand der Langzeitarchivierung von Digitalisaten: Die Ergebnisse der Fragebogenaktion*. Vortrag beim Expertengespräch: „Vom Digitalisierungsprojekt zur organisierten Langzeitarchivierung" am 11. November 2008 in München. http://www.digitale-sammlungen.de/mdz/content/veranstaltung/2008-11-11/02_expertengespraech_20081111_ergebnisse_fragebogen.pdf [20.05.09]

Lyon, Liz (2007): *Dealing with Data: Roles, Rights, Responsibilities and Relationships, consultancy report*. UKOLN, Bath, 2007. http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf [20.05.09]

Metje, Daniel (2008): *Organisations- und Geschäftsmodelle*. Vortrag beim KoLaWiss-Expertenworkshop, 9. Dezember 2008, Göttingen. http://kolawiss.uni-goettingen.de/workshop/AP6Orga.pdf [20.05.09]

Metje, Daniel / Vlaeminck, Sven (2008): Die Vision einer Kooperation von Forschenden und Daten-Kuratoren zur Sicherung unseres digitalen wissenschaftlichen Erbes. In: Ockenfeld, M.: *Verfügbarkeit von Informationen*. 30. Online-Tagung der DGI, 60. Jahrestagung der DGI, 14. bis 17. Oktober 2008, Frankfurt (Main), 2008. S. 107-117.

nestor-Arbeitsgruppe Vertrauenswürdige Archive – Zertifizierung (2008): *nestor: Kriterien - Kriterienkatalog vertrauenswürdige digitale Langzeitarchive. Version 2.* nestor Materialien 8, 2008. http://nbn-resolving.de/urn:nbn:de:0008-2008021802 [20.05.09]

Online Computer Library Center (OCLC) (2006): *OCLC Digital Archive Preservation Policy and Supporting Documentation*. Dublin, Ohio, USA. 2006. http://www.oclc.org/support/documentation/digitalarchive/preservationpolicy.pdf [20.05.09]

Organisation for Economic Co-operation and Development (OECD) (2004): *Science, Technology and Innovation for the 21st Century*. Meeting of the OECD Committee for Scientific and Technological Policy at Ministerial Level, 29-30 Januar 2004 - Final Communiqué, Paris, Frankreich, 2004. http://www.oecd.org/document/0,2340,en_2649_34487_25998799_1_1_1_1,00.html [20.05.09]

Organisation for Economic Co-operation and Development (OECD) (2006): *Recommendation of the Council concerning Access to Research Data from Public Funding*. Paris, Frankreich, 2006.

http://webdomino1.oecd.org/horizontal/oecdacts.nsf/Display/3A5FB1397B5ADFB7C12572980053C9D3?OpenDocument [20.05.09]

Organisation for Economic Co-operation and Development (OECD) (2007): *OECD Principles and Guidelines for Access to Research Data from Public Funding, Organisation for Economic Co-operation and Development*. Paris, Frankreich, 2007. http://www.oecd.org/dataoecd/9/61/38500813.pdf [20.05.09]

PARSE.Insight Projekt (2009): *Inventory of communities. Deliverable 3.2.* 2009. http://www.parse-insight.eu/downloads/PARSE-Insight_D3-2_InventoryOfCommunities_final.pdf [20.05.09]

Pfeiffenberger, Hans (2007): Offener Zugang zu wissenschaftlichen Primärdaten. In: *Zeitschrift für Bibliothekswesen und Bibliographie*. Heft 4-5. 2007 (54. Jg.). S. 207-210.

Projekt kopal (2007): *kopal: Ein Service für die Langzeitarchivierung digitaler Informationen.* http://kopal.langzeitarchivierung.de/downloads/kopal_Services_2007.pdf [20.05.09]

Research Information Network (RIN) (2008): *To Share or not to Share: Publication and Quality Assurance of Research Data Outputs*. Report commissioned by the Research Information Network. London, 2008. http://www.rin.ac.uk/files/Data%20publication%20report,%20main%20-%20final.pdf [20.05.09]

Scheer, Christian / Deelmann, Thomas / Loos, Peter (2003): *Geschäftsmodelle und internetbasierte Geschäftsmodelle – Begriffsbestimmung und Teilnehmermodell*. Mainz, 2003. http://wi.bwl.uni-mainz.de/publikationen/isym012.pdf [20.05.09]

Severiens, Thomas / Hilf, Eberhard R. (2006): *Studie zum Stand vorhandener Forschungsdaten und Rohdaten aus wissenschaftlichen Tätigkeiten: Erfordernisse und Eignung zur Archivierung bzw. Zurverfügungstellung in Deutschland (Primärdaten)*. nestor Materialien 6. http://nbn-resolving.de/urn:nbn:de:0008-20051114018 [20.05.09]

TIB Hannover (2009): *Europäische Initiative erleichtert den Zugang zu Forschungsdaten.* 2009. http://www.tib-hannover.de/de/die-tib/aktuelles/aktuelles/id/114/ [20.05.09]

Vögler, Max (2007): Die Deutsche Forschungsgemeinschaft und deutsche Initiativen zur Langzeitarchivierung für die Wissenschaft. In: *Geschichte im Netz : Praxis, Chancen, Visionen; Beiträge der Tagung .hist2006*.

Berlin, Clio-online und Humboldt-Universität zu Berlin, 2007, S. 244 – 257 (Historisches Forum 10 (2007), Teilband I) http://edoc.hu-berlin.de/histfor/10_I/PDF/Langzeitarchivierung_2007-10-I.pdf [20.05.09]

Wollenschläger, Thomas (2008): Geschäftsmodelle. In: Neuroth, Heike et al. (Hrsg.): *nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Version 1.5. Göttingen, 2008, S.5-1 bis 5-12. (2008). http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_224.pdf [20.05.09]